PCT

LD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



(51) International Patent Classification 6:		(1	1) International Publication Number:	WO 99/19493
C12N 15/53, 15/82, 9/02, 5/00, A01H 5/00, 5/10	A2	(4	3) International Publication Date:	22 April 1999 (22.04.99)
(21) International Application Number: PCT/U	S98/208	07	(81) Designated States: AL, AM, AT, A BA, BB, BG, BR, BY, CA, CH	
(22) International Filing Date: 5 October 1998	(05.10.9	98)	model), DE, DE (Utility model) EE, EE (Utility model), ES, FI, I GH, GM, HR, HU, ID, IL, IS,), DK, DK (Utility model), FI (Utility model), GB, GE,
(30) Pricrity Data: 08/948,564 10 October 1997 (10.10.97	') (US	LC, LK, LR, LS, LT, LU, LV, MX, NO, NZ, PL, PT, RO, RU	MD, MG, MK, MN, MW,

(71) Applicant (for all designated States except US): NORTH CAROLINA STATE UNIVERSITY [US/US]; 103 Bynum Hall, CB #7003, Raleigh, NC 27695–7003 (US).

(72) Inventors; and

- (75) Inventors/Applicants (for US only): SIMINSZKY, Balazs [HU/US]; 708 F Chapel Hill Drive, Raleigh, NC 27606 (US). DEWEY, Ralph, E. [US/US]; 9432 Cartersville Court, Raleigh, NC 27613 (US). CORBIN, Frederick, T. [US/US]; 4508 Lead Mine Road, Raleigh, NC 27612 (US).
- (74) Agents: BENNETT, Virginia, C. et al.; Myers, Bigel, Sibley & Sajovec, P.A., P.O. Box 37428, Raleigh, NC 27627 (US).

81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: CYTOCHROME P-450 CONSTRUCTS AND METHODS OF PRODUCING HERBICIDE-RESISTANT TRANSGENIC PLANTS

(57) Abstract

DNA sequence encoding novel cytochrome P-450 molecules are provided. The use of DNA constructs containing such molecules to transform plants is described, as are transgenic plants exhibiting increased resistance to phenylurea herbicides. Methods of using such DNA constructs and transformed plants are provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
\mathbf{AU}	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Монасо	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	ТJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	MŁ	Mali	TТ	Trinidad and Tobago
Bj	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	1T	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JР	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PI.	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

10

15

20

25

NOVEL CYTOCHROME P-450 CONSTRUCTS AND METHODS OF PRODUCING HERBICIDE-RESISTANT TRANSGENIC PLANTS

Field of the Invention

The present invention relates to DNA encoding novel cytochrome P-450 molecules, and the transformation of cells with such DNA. These DNA sequences may be used in methods of producing plants with an altered ability to metabolize chemical compounds, such as phenylurea herbicides.

Background of the Invention

Cytochrome P-450 (P-450) monooxygenases are ubiquitous hemoproteins present in microorganisms, plants and animals. Comprised of a large and diverse group of isozymes, P-450s mediate a great array of oxidative reactions using a wide range of compounds as substrates, and including biosynthetic processes such as phenylpropanoid, fatty acid, and terpenoid biosynthesis; metabolism of natural products; and detoxification of foreign substances (xenobiotics). *See e.g.*, Schuler, *Crit. Rev. Plant Sci.* 15:235-284 (1996). In a typical P-450 catalyzed reaction, one atom of molecular oxygen (O₂) is incorporated into the substrate, and the other atom is reduced to water by NADPH. For most-eucaryotic P-450s, NADPH:cytochrome P-450 reductase, a membrane-bound flavoprotein, transfers the necessary two electrons from NADPH to the P-450 (Bolwell et al, *Phytochemistry* 37: 1491-1506 (1994)).

Frear et al. (Phytochemistry 8:2157-2169 (1969)) demonstrated the metabolism of monuron by a mixed-function oxidase located in a microsomal fraction of cotton seedlings. Further evidence has accumulated supporting the involvement of P-450s in the metabolism and detoxification of numerous herbicides representing several distinct classes of compounds (reviewed in Bolwell et al., 1994; Schuler, 1996). Differential herbicide metabolizing P-450 activities are believed to represent one of the mechanisms that enables certain crop species to be more tolerant of a particular herbicide than other crop or weedy species.

10

15

20

25

30

Summary of the Invention

A first aspect of the present invention is an isolated DNA molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15 or SEQ ID NO:17; or DNA sequences which encode an enzyme of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18; or DNA sequences which have at least about 90% sequence identity to the above DNA and which encode a cytochrome P450 enzyme; and DNA sequences which differ from the above DNA due to the degeneracy of the genetic code.

A further aspect of the present invention is a cytochrome p450 enzyme having an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6; SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18.

A further aspect of the present invention is an isolated DNA molecule comprising SEQ ID NO:1; DNA sequences which encode an enzyme of SEQ ID NO:2,; DNA sequences which have at least about 90% sequence identity to the above DNA and which encode a cytochrome P450 enzyme; and DNA sequences which differ from the above DNA due to the degeneracy of the genetic code.

A further aspect of the present invention is a cytochrome p450 peptide of SEQ ID NO:2.

A further aspect of the present invention is a DNA construct comprising a promoter operable in a plant cell and a DNA segment encoding a peptide of SEQ ID NO:2 downstream from and operatively associated with the promoter.

A further aspect of the present invention is a method of making a transgenic plant cell having an increased ability to metabolize phenylurea compounds compared to an untransformed plant cell. The plant cell is transformed with an exogenous DNA construct comprising a promoter operable in a plant cell and a DNA sequence encoding a peptide of SEQ ID NO:2. Transformed plants, seed and progeny of such plants are also aspects of the

present invention.

A further aspect of the present invention is a transgenic plant having an increased ability to metabolize phenylurea compounds. Such transgenic plants contain exogenous DNA encoding a peptide of SEQ ID NO:2.

5

10

15

20

25

30

Brief Description of the Drawings

Figure 1 depicts dithionite-reduced carbon monoxide difference spectra, where the solid line represents microsomes isolated from yeast transformed with CYP71A10, and the dotted line shows the difference spectra from yeast transformed with control vector V-60. Microsomal protein concentration was 1 mg/ml.

Figure 2 shows thin-layer chromatograms of [14 C]-radiolabeled fluometuron, linuron, chlortoluron, and diuron and their respective metabolites after incubation of the radiolabeled herbicides with yeast microsomes containing the CYP71A10 protein. Initial substrate concentrations for fluometuron, linuron, chlortoluron and diuron were 5.2, 6.5, 4.0, and 3.7 μ M, respectively. P = parent compound; M = metabolite.

Figure 3 shows the chemical structures of fluometuron, linuron, chlortoluron and diuron, and their previously characterized metabolites. The linuron and chlortoluron metabolites are designated major or minor depending on their predicted relative abundance in assays using yeast microsomes containing the soybean CYP71A10 protein.

Figure 4 shows thin-layer chromatograms using [¹⁴C]-radiolabeled linuron in various control reactions. The complete reaction mixture (COMPLETE) contained 3.2 μM linuron, 0.75 mM NADPH and 2.5 mg/ml microsomal protein isolated from CYP71A10-transformed yeast in 50 mM phosphate buffer (pH 7.1). Other reactions varied from COMPLETE by the addition of carbon monoxide (+CO), the omission of NADPH (NO NADPH), or the use of yeast microsomes isolated from cells expressing the control vector (V-60). P = parent compound; M = metabolite.

10

15

20

25

30

Figure 5A shows tobacco line 25/2 plants (transformed with soybean CYP71A10) grown on media containing no herbicide.

Figure 5B shows control tobacco plants (transformed with vector pBI121) grown on media containing $0.5~\mu M$ linuron.

Figure 5C shows tobacco line 25/2 (transformed with soybean CYP71A10) individuals grown on media containing 0.5 μ M linuron.

Figure 5D shows tobacco line 25/2 (transformed with soybean CYP71A10) individuals grown on media containing 2.5 μ M linuron.

Figure 5E shows control tobacco plants (transformed with vector pBI121) grown on media containing 1.0 μ M chlortoluron.

Figure 5F shows tobacco line 25/2 (transformed with soybean CYP71A10) individuals grown on media containing 1.0 μ M chlortoluron.

Detailed Description of the Invention

1. Overview of the present research:

The present inventors utilized a strategy based on the random isolation and screening of soybean cDNAs encoding cytochrome P-450 (P-450) isozymes to identify P-450 isozymes involved in herbicide metabolism. Eight full-length and one near full-length P-450 cDNAs representing eight distinct P-450 families were isolated using polymerase chain reaction (PCR)-based technologies (SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15 and 17). Five of these soybean P-450 cDNAs were successfully overexpressed in yeast, and microsomal fractions generated from these strains were tested for their potential to mediate the metabolism of ten herbicides and one insecticide. *In vitro* enzyme assays showed that the gene product of one heterologously expressed P-450 cDNA (CYP71A10) (SEQ ID NO:1) specifically mediated the metabolism of phenylurea herbicides, converting four herbicides of this class (fluometuron, linuron, chlortoluron, and diuron) into more polar metabolites. Analyses of the metabolites indicate that the CYP71A10 encoded enzyme functions primarily as an N-demethylase with regard to

-5-

fluometuron, linuron and diuron, and as a ring-methyl hydroxylase when chlortoluron is the substrate. *In vivo* assays using excised leaves demonstrated that all four herbicides were more readily metabolized in CYP71A10-transformed tobacco in comparison to centrol plants.

Shiota et al. reported that fused constructs derived from the rat CYP1A1 and yeast NADPH-cytochrome P-450 oxidoreductase cDNAs conferred chlortoluron resistance in tobacco by enhancing herbicide metabolism (Shiota et al., Plant Physiol. 106:17-23 (1994)). In another study, a chloroplast-targeted. bacterial CYP105A1 expressed in tobacco catalyzed the toxification of R7402, a sulfonylurea pro-herbicide (O'Keefe et al., Plant Physiol. 105:473-482 (1994)). The cloning and heterologous expression of an endogenous plant P-450 gene that is potentially involved in herbicide metabolism was reported by Pierrel et al., Eur. J. Biochem. 224:835-844 (1994), where a trans-cinnamic acid hydroxylase cDNA (CYP73A1) isolated from artichoke and expressed in yeast catalyzed the ring-methyl hydroxylation of chlortoluron. In vivo experiments with artichoke tubers, however, demonstrated that the ring-methyl hydroxy metabolite represented only a minor portion of the metabolites produced and that the major metabolite was demethylated chlortoluron (Pierrel et al., 1994). This together with the observation that the turnover number of the heterologously expressed enzyme was very low (0.014/ min), suggested that CYP73A1 plays a minimal role in chlortoluron metabolism in vivo. US Patent No. 5,349,127 to Dean et al. discloses the use of DNA encoding certain P-450 enzymes, isolated from Streptomyces griseolus, to produce transformed plants with increased metabolism of certain compounds. (All US patents referred to herein are intended to be incorporated herein in their entirety.)

Although the role of P-450 enzymes in catalyzing the metabolism of a variety of herbicides has been documented, little progress has been made in the identification of the endogenous plant P-450s that are responsible for degrading these compounds. Protein purification of specific isozymes involved in the metabolism of a specific herbicide has been hindered by the instability of the

5

10

15

20

25

enzymes, their low concentrations in most plant tissues, and difficulties in the reconstitution of active complexes from solubilized components. Furthermore, any given plant tissue may possess dozens, if not hundreds, of unique P-450 isozymes, complicating the purification to homogeneity of a particular isozyme. Because plants have only been exposed to phenylurea herbicides during the past few decades, it is unlikely that enzymes have evolved solely for the purposed of metabolizing this class of xenobiotics.

2. Use of CYP71A10 to produce phenylurea-resistant plants:

The present invention provides materials and methods useful in producing transgenic plant cells and plants with increased resistance to phenylurea herbicides. Increased herbicide resistance, as used herein, refers to the ability of a plant to withstand levels of an herbicide that have a negative impact on wildtype (untransformed) plants of the same species and/or variety. Resistance, as used herein, does not necessarily mean that the resistant plant is completely unaffected by exposure to the herbicide; rather, resistant plants suffer less extensive or less severe damage than comparable wild-type plants. Methods of assessing the extent and/or severity of herbicide impact will vary depending on the particular plant and the particular herbicide being tested; such assessment methods will be apparent to those skilled in the art. The negative effects of a herbicide may be evidenced by the complete arrest of plant growth, or by an inhibition in the rate or amount of growth. Additionally, methods of the present invention may be used to decrease herbicide residues in plants, even where the amounts of herbicides present in the plant do not cause an appreciable negative effect on the plant as a whole.

Increased resistance to a herbicide can be due to an increased ability to metabolize a herbicide to less harmful metabolites. Accordingly, plants of the present invention which exhibit increased resistance to a herbicide may also be described as having an increased ability to metabolize the starting herbicidal compound, where the metabolites are less harmful to the plant than the starting

5

10

15

20

25

10

15

20

25

30

compound.

In the examples provided herein, yeast microsomes and transgenic tobacco plants expressing the CYP71A10 peptide (SEQ ID NO:2) and exposed to various phenylurea herbicides produced the same degradation products that have previously been observed when these same compounds have been incubated with metabolically active plant microsomes. These results indicate that the CYP71A10 peptide plays a role in the effective metabolism of phenylurea herbicides.

The present examples demonstrate that the overexpression of a CYP71A10 peptide of SEQ ID NO:2 in tobacco enhanced the plant's capacity to metabolize all four phenylurea herbicides tested, and that appreciable levels of tolerance were conferred to linuron and chlortoluron. Fluometuron was the most actively metabolized compound in both the yeast and transgenic plant systems, yet the enhancement in tolerance to this herbicide at the whole plant level was not as great as for linuron and chlortoluron. While not wishing to be held to a single theory, the present inventors surmise that the lack of correlation between the rate of herbicide metabolism and herbicide tolerance may be explained by the differential toxicities of the various phenylurea derivatives produced in the Consistent with this hypothesis are the CYP71A10-transformed tobacco. previous observations that N-demethyl derivatives of fluometuron, diuron and chlortoluron are only moderately less toxic than their parent compounds (Rubin and Eshel, Weed Sci. 19:592-594 (1971); Dalton et al., Weeds 14:31-33 (1966); Ryan and Owen, Proc. Brit. Crop Prot. Conf. Weeds 1:317-324 (1982)). In contrast, linuron is a 10-fold greater inhibitor of the Hill-reaction than Ndemethyl linuron (Suzuki and Casida, J. Agric. Food Chem. 29:1027-1033 (1981)), and the hydroxylated and the didemethlayed derivatives of chlortoluron are considered to be nonherbicidal (Ryan and Owen, 1982).

The present inventors found that the relative rates of herbicide metabolism in leaves of CYP71A10-transformed tobacco and in yeast microsomes assayed *in vitro* were similar (see **Tables 4** and 5). With the exception of the transgenic

10

15

20

25

30

plant leaves showing a somewhat greater metabolic activity against chlortoluron than was apparent in the yeast microsomal assays, both systems followed the general order of metabolism of fluometuron \geq linuron > chlortoluron > diuron. These results indicate that expression of a test plant P-450 in yeast and quantification of the metabolism of a test compound using yeast microsomes, is a suitable system for screening plant P-450s for their metabolic function, and for their potential usefulness in the production of transgenic plants with altered metabolism of chemical compounds such as herbicides and insecticides.

The present inventors have shown that the random isolation of P-450 cDNAs with subsequent heterologous expression in yeast is an effective strategy to characterize cDNAs whose product is capable of affecting the metabolism of a test compound. This approach is useful in characterizing the substrates (both natural and artificial) affected by a P-450, in determining the function of P-450 genes whose catalytic activities remain unclear, and in screening P-450s for the ability to increase or decrease the metabolism of a test compound. A particularly useful aspect of this method is the ability to screen isolated P-450s for their effects on the metabolism by plants of herbicides, insecticides, or other chemical compounds. Increased metabolism may result in enhanced resistance to the effects of a compound (where the metabolites are less harmful than the starting compound), or in increased sensitivity to the effects of a compound (where one or more metabolites are more toxic than the starting compound; see O'Keefe et al., 1994).

3. DNA Constructs:

Those familiar with recombinant DNA methods available in the art will recognize that one can employ a cDNA molecule (or a chromosomal gene or genomic sequence) encoding a P-450 peptide, joined in the sense orientation with appropriate operably linked regulatory sequences, to construct transgenic cells and plants. (Those of skill in the art will also recognize that appropriate regulatory sequences for expression of genes in the sense orientation include any

-9-

one of the known eukaryotic translation start sequences, in addition to the promoter and polyadenylation/transcription termination sequences described herein). Appropriate selection of the encoded P-450 peptide will provide transformed plants characterized by altered (enhanced or retarded) metabolism of phenylurea compounds.

DNA constructs, or "transcription cassettes," of the present invention include, 5' to 3' in the direction of transcription, a promoter as discussed herein, a DNA sequence as discussed herein operatively associated with the promoter, and, optionally, a termination sequence including stop signal for RNA polymerase and a polyadenylation signal for polyadenylase. All of these regulatory regions should be capable of operating in the cells of the tissue to be transformed. Any suitable termination signal may be employed in carrying out the present invention, examples thereof including, but not limited to, the nopaline synthase (nos) terminator, the octapine synthase (ocs) terminator, the CaMV terminator, or native termination signals derived from the same gene as the transcriptional initiation region or derived from a different gene. See, e.g., Rezian et al. (1988) supra, and Rodermel et al. (1988), supra.

The term "operatively associated," as used herein, refers to DNA sequences on a single DNA molecule which are associated so that the function of one is affected by the other. Thus, a promoter is operatively associated with a DNA when it is capable of affecting the transcription of that DNA (i.e., the DNA is under the transcriptional control of the promoter). The promoter is said to be "upstream" from the DNA, which is in turn said to be "downstream" from the promoter.

The transcription cassette may be provided in a DNA construct which also has at least one replication system. For convenience, it is common to have a replication system functional in Escherichia coli, such as ColE1, pSC101, pACYC184, or the like. In this manner, at each stage after each manipulation, the resulting construct may be cloned, sequenced, and the correctness of the manipulation determined. In addition, or in place of the E. coli replication

5

10

15

20

25

-10-

system, a broad host range replication system may be employed, such as the replication systems of the P-1 incompatibility plasmids, e.g., pRK290. In addition to the replication system, there will frequently be at least one marker present, which may be useful in one or more nosts, or different markers for individual hosts. That is, one marker may be employed for selection in a prokaryotic host, while another marker may be employed for selection in a eukaryotic host, particularly the plant host. The markers may be protection against a biccide, such as antibiotics, toxins, heavy metals, or the like; may provide complementation, by imparting prototrophy to an auxotrophic host; or may provide a visible phenotype through the production of a novel compound in the plant.

The various fragments comprising the various constructs, transcription cassettes, markers, and the like may be introduced consecutively by restriction enzyme cleavage of an appropriate replication system, and insertion of the particular construct or fragment into the available site. After ligation and cloning the DNA construct may be isolated for further manipulation. All of these techniques are amply exemplified in the literature as exemplified by J. Sambrook et al., Molecular Cloning, A Laboratory Manual (2d Ed. 1989)(Cold Spring Harbor Laboratory).

Vectors which may be used to transform plant tissue with nucleic acid constructs of the present invention include both Agrobacterium vectors and ballistic vectors, as well as vectors suitable for DNA-mediated transformation.

4. Promoters:

5

10

15

20

The term 'promoter' refers to a region of a DNA sequence that incorporates the necessary signals for the efficient expression of a coding sequence. This may include sequences to which an RNA polymerase binds but is not limited to such sequences and may include regions to which other regulatory proteins bind together with regions involved in the control of protein translation and may include coding sequences.

-11-

Promoters employed in carrying out the present invention may be constitutively active promoters. Numerous constitutively active promoters which are operable in plants are available. A preferred example is the Cauliflower Mosaic Virus (CaMV) 35S promoter which is expressed constitutively in most plant tissues. Use of the CaMV promoter for expression of recombinant genes in tobacco roots has been well described (Lam et al., "Site-Specific Mutations Alter In Vitro Factor Binding and Change Promoter Expression Pattern in Transgenic Plants", Proc. Nat. Acad. Sci. USA 86, pp. 7890-94 (1989); Poulsen et al. "Dissection of 5' Upstream Sequences for Selective Expression of the Nicotiana plumbaginifolia rbcS-8B Gene", Mol. Gen. Genet. 214, pp. 16-23 (1988)). In the alternative, the promoter may be a tissue-specific promoter or a promoter that is expressed temporally or developmentally. See, e.g., US Patent No. 5,459,252 to Conkling et al.; Yamamoto et al., The Plant Cell, 3:371 (1991). In methods of transforming plants to alter the effects of herbicides or to decrease residual amounts of herbicides or pesticides in plants, selection of a suitable promoter will vary depending on the plant species, the specific chemical compound used as a herbicide or pesticide, and the time and method of applying the chemical compound to the plant or plant crop, as will be apparent to those skilled in the art.

20

25

30

5

10

15

5. Selectable Markers:

The recombinant DNA molecules and vectors used to produce the transformed cells and plants of this invention may further comprise a dominant selectable marker gene. Suitable dominant selectable markers include, inter alia, antibiotic resistance genes encoding neomycin phosphotransferase (NPTII), hygromycin phosphotransferase (HPT), and chloramphenicol acetyltransferase (CAT). Another well-known dominant selectable marker suitable is a mutant dihydrofolate reductase gene that encodes methotrexate-resistant dihydrofolate reductase. DNA vectors containing suitable antibiotic resistance genes, and the corresponding antibiotics, are commercially available. Transformed cells are

selected out of the surrounding population of non-transformed cells by placing the mixed population of cells into a culture medium containing an appropriate concentration of the antibiotic (or other compound normally toxic to the untransformed cells) against which the chosen dominant selectable marker gene product confers resistance. Thus, only those cells that have been transformed will survive and multiply.

A further aspect of the present invention is use of the identified P-450 coding sequences as a selectable marker gene. A DNA construct comprising a sequence encoding a P-450 known to increase resistance to a compound (such as SEQ ID NO:2) is utilized to transform cells, in accordance with methods known in the art. Those cells that subsequently exhibit resistance to the compound are indicated as transformed. Such constructs may be used to verify the success of a transformation technique or to select transformed cells of interest.

15

20

25

30

10

5

6. Sequence similarity and hybridization conditions:

Nucleic acid sequences employed in carrying out the present invention include those with sequence similarity to SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17, and encoding a protein having P-450 enzymatic activity. This definition is intended to encompass natural allelic variants and minor sequence variations in the nucleic acid sequence encoding a P-450 molecule, or minor sequence variations in the amino acid sequence of the encoded product. Thus, DNA sequences that hybridize to DNA of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17 and code for expression of a P-450 enzyme, particularly a plant P-450 enzyme, may also be employed in carrying out aspects of the present invention. The nomenclature for P-450 genes is based on amino acid sequence identity; methods of determining sequence similarity are well-known to those skilled in the art. Typically, sequences sharing >40% identity are placed in the same family, >55% identity defines members of the same subfamily, and sequences that

display >97% identity are assumed to represent allelic variants. Conditions which permit other DNA sequences which code for expression of a protein having P-450 enzymatic activity to hybridize to DNA of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17, or to other DNA sequences encoding the protein given as SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16 or 18 can be determined in a routine manner. For example, hybridization of such sequences may be carried out under conditions of reduced stringency or even stringent conditions (e.g., conditions represented by a wash stringency of 0.3 M NaCl, 0.03 M sodium citrate, 0.1% SDS at 60°C or even 70°C to DNA encoding the protein given as SEQ ID NO:2 herein in a standard in situ hybridization assay. See J. Sambrook et al., Molecular Cloning, A Laboratory Manual (2d Ed. 1989)(Cold Spring Harbor Laboratory)). In general, such sequences will be at least 65% similar, 75% similar, 80% similar, 85% similar, 90% similar, 93% similar, 95% similar, or even 97% or 98% similar, or more, with the sequence given herein as SEQ ID NO:1, or DNA sequences encoding proteins of SEQ ID NO:2. (Determinations of sequence similarity are made with the two sequences aligned for maximum matching; gaps in either of the two sequences being matched are allowed in maximizing matching. Gap lengths of 10 or less are preferred, gap lengths of 5 or less are more preferred, and gap lengths of 2 or less still more preferred.)

As used herein, the term 'gene' refers to a DNA sequence that incorporates (1) upstream (5') regulatory signals including a promoter, (2) a coding region specifying the product, protein or RNA of the gene, (3) downstream (3') regions including transcription termination and polyadenylation signals and (4) associated sequences required for efficient and specific expression.

The DNA sequence of the present invention may consist essentially of a sequence provided herein (SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17), or equivalent nucleotide sequences representing alleles or polymorphic variants of these genes, or coding regions thereof.

30 Use of the phrase "substantial sequence similarity" in the present

5

10

15

20

-14-

specification and claims means that DNA, RNA or amino acid sequences which have slight and non-consequential sequence variations from the actual sequences disclosed and claimed herein are considered to be equivalent to the sequences of the present invention. In this regard, "slight and non-consequential sequence variations" mean that "similar" sequences (i.e., the sequences that have substantial sequence similarity with the DNA, RNA, or proteins disclosed and claimed herein) will be functionally equivalent to the sequences disclosed and claimed in the present invention. Functionally equivalent sequences will function in substantially the same manner to produce substantially the same compositions as the nucleic acid and amino acid compositions disclosed and claimed herein.

DNA sequences provided herein can be transformed into a variety of host cells. A variety of suitable host cells, having desirable growth and handling properties, are readily available in the art.

Use of the phrase "isolated" or "substantially pure" in the present specification and claims as a modifier of DNA, RNA, polypeptides or proteins means that the DNA, RNA, polypeptides or proteins so designated have been separated from their *in vivo* cellular environments through the efforts of human beings.

As used herein, a "native DNA sequence" or "natural DNA sequence" means a DNA sequence which can be isolated from non-transgenic cells or tissue. Native DNA sequences are those which have not been artificially altered, such as by site-directed mutagenesis. Once native DNA sequences are identified, DNA molecules having native DNA sequences may be chemically synthesized or produced using recombinant DNA procedures as are known in the art. As used herein, a native plant DNA sequence is that which can be isolated from non-transgenic plant cells or tissue.

7. Transformed plants:

Methods of making recombinant plants of the present invention, in general, involve first providing a plant cell capable of regeneration (the plant cell

5

10

15

20

-15-

typically residing in a tissue capable of regeneration). The plant cell is then transformed with a DNA construct comprising a transcription cassette of the present invention (as described herein) and a recombinant plant is regenerated from the transformed plant cell. As explained below, the transforming step is carried out by techniques as are known in the art, including but not limited to bombarding the plant cell with microparticles carrying the transcription cassette, infecting the cell with an Agrobacterium tumefaciens containing a Ti plasmid carrying the transcription cassette, or any other technique suitable for the production of a transgenic plant.

Numerous Agrobacterium vector systems useful in carrying out the present invention are known. For example, U.S. Patent No. 4,459,355 discloses a method for transforming susceptible plants, including dicots, with an Agrobacterium strain containing the Ti plasmid. The transformation of woody plants with an Agrobacterium vector is disclosed in U.S. Patent No. 4,795,855. Further, U.S. Patent No. 4,940,838 to Schilperoort et al. discloses a binary Agrobacterium vector (i.e., one in which the Agrobacterium contains one plasmid having the vir region of a Ti plasmid but no T region, and a second plasmid having a T region but no vir region) useful in carrying out the present invention.

Microparticles carrying a DNA construct of the present invention, which microparticle is suitable for the ballistic transformation of a plant cell, are also useful for making transformed plants of the present invention. The microparticle is propelled into a plant cell to produce a transformed plant cell, and a plant is regenerated from the transformed plant cell. Any suitable ballistic cell transformation methodology and apparatus can be used in practicing the present invention. Exemplary apparatus and procedures are disclosed in Sanford and Wolf, U.S. Patent No. 4,945,050, and in Christou et al., U.S. Patent No. 5,015,580. When using ballistic transformation procedures, the transcription cassette may be incorporated into a plasmid capable of replicating in or integrating into the cell to be transformed. Examples of microparticles suitable

5

10

15

20

25

-16-

for use in such systems include 1 to $5 \mu m$ gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

Plant species may be transformed with the DNA construct of the present invention by the DNA-mediated transformation of plant cell protoplasts and subsequent regeneration of the plant from the transformed protoplasts in accordance with procedures well known in the art. Fusion of tobacco protoplasts with DNA-containing liposomes or via electroporation is known in the art. (Shillito et al., "Direct Gene Transfer to Protoplasts of Dicotyledonous and Monocotyledonous Plants by a Number of Methods, Including Electroporation", Methods in Enzymology 153, pp. 313-36 (1987)).

As used herein, transformation refers to the introduction of exogenous DNA into cells, so as to produce transgenic cells stably transformed with the exogenous DNA. Transformed plant cells are induced to regenerate intact plants through application of cell and tissue culture techniques that are well known in the art. The method of plant regeneration is chosen so as to be compatible with the method of transformation. The stable presence and the orientation of the exogenous DNA in transgenic plants can be verified by Mendelian inheritance of the DNA sequence, as revealed by standard methods of DNA analysis applied to progeny resulting from controlled crosses.

Plants of horticultural or agronomic utility, such as vegetable or other crops, can be transformed according to the present invention using techniques available in the art. A plant suitable for use in the present methods is Nicotiana tabacum, or tobacco. Any strain or variety of tobacco may be used.

25 Additional plants (both monocots and dicots) which may be employed in practicing the present invention include, but are not limited to, potato (Solanum tuberosum), soybean (Glycine max), tomato (Lycopersicon esculentum), peanuts (Arachis hypogaea), cotton (Gossypium hirsutum), green beans (Phaseolus vulgaris), lima beans (Phaseolus limensis), peas (Lathyrus spp.)cassava (Manihot esculenta), coffee (Cofea spp.), pineapple (Ananas comosus), citrus trees (Citrus

5

10

15

-17-

spp.), banana (Musa spp.), corn (Zea mays), oilseed rape (Brassica napus), wheat, oats, barley, rye and rice. Thus, an illustrative category of plants which may be used to practice aspects of the present invention are the dicots, and a more particular category of plants which may be used to practice the present invention are members of the family Solanacae.

The methods of the present invention can further be practiced with turfgrass, including cool season turfgrasses and warm season turfgrasses. Examples of cool season turfgrasses are Bluegrasses (Poa L.), such as Kentucky Bluegrass (Poa pratensis L.), rough Bluegrass (Poa trivialis L.), Canada Bluegrass (Poa compressa L.), Annual Bluegrass (Poa annua L.), Upland 10 Bluegrass (Poa glaucantha Gaudin), Wood Bluegrass (Poa nemoralis L.), and Bulbous Bluegrass (Poa bulbosa L.); the Bentgrasses and Redtop (Agrostis L.), such as Creeping Bentgrass (Agrostis palustris Huds.), Colonial Bentgrass (Agrostis tenius Sibth.), Velvet Bentgrass (Agrostis canina L.), South German. Mixed Bentgrass (Agrostis L.), and Redtop (Agrostis alba L.); the Fescues 15 (Festuca L.), such as Red Fescue (Festuca rubra L.), Chewings Fescue (Festuca rubra var. commutata Gaud.), Sheep Fescue (Festuca ovina L.), Hard Fescue (Festuca ovina var. duriuscula L. Koch), Hair Fescue (Festuca capillata Lam.), Tall Fescue (Festuca arundinacea Schreb.), Meadow Fescue (Festuca elatior L.); the Rye grasses (Lolium L.), such as Perennial Ryegrass (Lolium perenne L.), 20 Italian Ryegrass (Lolium multiflorum Lam.); the Wheatgrasses (Agropyron Gaertn.), such as Fairway Wheatgrass (Agropyron cristatum L. Gaertn.), Western Wheatgrass (Agropyron smithii Rydb.). Examples of warm season turfgrasses are the Bermudagrasses (Cynodon L.C. Rich), the Zoysiagrasses (Zoysia Willd.), St. Augustinegrasses (Stenotaphrum secundatum (Walt.) 25 Kuntze), Centipedegrass (Eremochioa ophiuroides (Munro.) Hack.), Carpetgrass (Axonopus Beauv.), Bahiagrass (Paspalum notatum Flugge.), Kikuyugrass (Pennisetum clandestinum Hochst. ex Chiov.), Buffalograss (Buchloe dactyloides (Nutt.) Engelm.), Blue Grama (Bouteloua gracilis (H.B.K.) Lag. ex Steud.),

Sideoats Grama (Bouteloua curtipendula (Michx.) Torr.), and Dichondra

30

-18-

(Dichondra Forst.).

5

10

15

20

25

30

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a vector of the present invention. The term "organogenesis," as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers; the term "embryogenesis," as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

Plants of the present invention may take a variety of forms. The plants may be chimeras of transformed cells and non-transformed cells; the plants may be clonal transformants (e.g., all cells transformed to contain the transcription cassette); the plants may comprise grafts of transformed and untransformed tissues (e.g., a transformed root stock grafted to an untransformed scion in citrus species). The transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, first generation (or T1) transformed plants may be selfed to provide homozygous second generation (or T2) transformed plants, and the T2 plants further propagated through classical breeding techniques. A dominant selectable marker (such as nptII) can be associated with the transcription cassette to assist in breeding.

As used herein, a crop comprises a plurality of plants of the same genus or species, planted together in an agricultural field. By "agricultural field" is meant a common plot of soil or a greenhouse. Thus, the present invention provides a method of producing a crop of plants having altered metabolism of chemical compounds (such as a phenylurea herbicide), and thus having altered

-19-

resistance to the chemical compound, compared to a crop of non-transformed plants of the same genus or species, or variety.

Where a crop comprises a plurality of transgenic plants with increased resistance to phenylurea compounds according to the present invention, such compounds may be used as post-emergent herbicides to control undesirable plant species. Accordingly, a method of using phenylurea compounds as post-emergent herbicides according to the present invention comprises planting a plurality of transformed plant seed (or transformed plants) with enhanced resistance to a phenylurea herbicide, and applying that herbicide to the field after the germination and emergence of at least some of said transformed plant seed (or following the planting of transformed plants). Application of the phenylurea herbicide will selectively impact non-resistant plants.

9. Microbial decontamination:

5

10

15

20

25

30

Microbial cells useful for degrading phenylurea compounds, which cells contain and express a heterologous DNA molecule encoding a P-450 enzyme that enhances the metabolism of the phenylurea compound in the microbial cell (e.g., a peptide of SEQ ID NO:2), are a further aspect of the present invention. Suitable host microbial cells include soil microbes (i.e., those which grow in the soil) transformed to express a P-450 enzyme that enhances the metabolism of one or more phenylurea compounds by the host cell. Suitable microbes include bacteria (such as Agrobacterium, Bacillus, Streptomyces, Nocardia, etc.), fungi (including yeasts), and algae. Microbes can be selected, by methods known in the art of soil microbiology, to correspond to those which are typically found in the substrate to be treated. Liquids which are contaminated with phenylurea compounds may be contacted to transformed microorganisms by passing the contaminated liquid through a bioreactor which contains the microorganism. Numerous suitable bioreactor designs are known in the art. A microbial host particularly suitable for bioreactors is yeast.

Combination treatments utilizing aspects of the present invention involve

-20-

the application of a phenylurea compound in a location such as an agricultural field (e.g., as a herbicide), and subsequent application of a transformed microbe as described above in an amount effective to degrade residual applied herbicide. Application of the herbicide may be carried out in accordance with known techniques.

The examples which follow are set forth to illustrate the present invention, and are not to be construed as limiting thereof.

EXAMPLE 1

Materials and Methods

a. Substrates

Phenyl-U-[¹⁴C] fluometuron, phenyl-U-[¹⁴C] chlortoluron, phenyl-U-[¹⁴C] metolachlor, phenyl-U-[¹⁴C] prosulfuron, pyrimidinyl-2- diazinon, and phenyl-U-[¹⁴C] alachlor were provided by Novartis (Greensboro, North Carolina); phenyl-U-[¹⁴C] bentazon was donated by BASF (Research Triangle Park, North Carolina); phenyl-U-[¹⁴C] linuron, phenyl-U-[¹⁴C] diuron, and carbonyl-[¹⁴C] metribuzin were a gift from DuPont de Nemours (Wilmington, Delaware); carboxyl-[¹⁴C] imazaquin was provided by American Cyanamid (Princeton, New Jersey).

20

25

30

5

10

15

b. Isolation of P-450 cDNAs

Random amplification of partial cDNAs encoding P-450 enzymes was conducted essentially as described by Meijer et al., *Plant Mol. Biol.* 22:379-383 (1993), using a soybean (*Glycine max* cv Dare) leaf cDNA library as the template (Dewey et al., *Plant Cell* 6:1495-1507 (1994)). Briefly, degenerate inosine-containing primers were synthesized based on the highly conserved heme-binding region. The precise sequences of these primers are described in Meijer et al. (1993). An oligo-dT primer complementary to the poly(A) tail of the cDNA clones was used in conjunction with the degenerate primers in PCR amplification assays. Amplification products were cloned into the T-tailed pCRII plasmid

10

15

20

25

30

(Invitrogen, San Diego, CA) and DNA sequence analysis of the first 300-400 base pairs downstream of the conserved region was used to establish whether a given amplification product represented a true P-450 cDNA.

To recover full-length versions of the partial cDNAs, a primer (5'-TGTCTAACTCCTTCCTTTC-3') (SEQ ID NO:19) complementary to the pYES2 vector (the vector into which the soybean cDNA library was cloned) and a downstream primer corresponding to a segment of the 3' untranslated region for each of the unique P-450 cDNAs were used in PCR reactions using the same soybean cDNA library as the template. PCR products were again cloned into the pCRII plasmid and the entire DNA sequence was determined for the largest cDNA amplified for each unique soybean P-450.

To isolate full-length versions of the respective P-450 ORFs without including any of the 5' untranslated region (which has been shown to potentially impede gene expression in yeast (Pompon, Eur. J. Biochem. 177:285-293 (1988)), an additional PCR reaction was performed with two gene-specific primers. The forward primers contained a BamHI restriction site immediately followed by the ATG start codon, and the next 14-15 bases of the reading frame; the downstream primer was again specific for the 3' untranslated regions of the respective genes and included sequences specifying either EcoRI, KpnI, and SacI to facilitate subcloning of the P-450 cDNAs into the yeast expression vector, pYeDP60 (V-60; Urban et al., Biochimie 72:463-472 (1990)).

All PCR reactions, with the exception of the initial amplification of the partial P-450 cDNAs (see Meijer et al. (1993)), contained 0.2 ng/μl template, 2 μM of each primer, 200 μM of each dNTP, and 1.5 mM MgCl₂ in a final reaction volume of 50 μl. Amplification was initiated by the addition of 1.5 U EXPANDTM High Fidelity enzyme mix using conditions described by the manufacturer (Boeringer Mannheim). DNA sequence was determined by the chain termination method (Sanger et al., *Proc. Natl. Acad. Sci. USA* 74:5463-5467 (1977)) using fluorescent dyes (Applied Biosystems, Foster City, CA). DNA and predicted amino acid sequences were analyzed using the BLAST

algorithm and the GAP program (University of Wisconsin, Madison, Genetics Computing Group software package).

c. P-450 cDNA Expression in Yeast

5

10

15

20

25

30

Yeast transformation was performed as described by Geitz et al., *Nucleic Acids Research* 20:1425 (1992). Media composition, culturing conditions, galactose induction, and microsomal preparations were conducted according to Pompon et al., *Methods Enzymol*. 272:51-64 (1995), using a culture volume of 250 ml. Microsomal protein was quantified spectrophotometrically using the method of Waddell, *J. Lab. Clin. Med*. 48:311-314 (1956), using bovine albumin as a standard. Dithionite-reduced, carbon monoxide difference spectra was obtained as previously outlined (Estabrook and Werringloer, *Methods Enzymol*. 52:212-220 (1978)) using a Shimadzu Recording Spectrophotometer UV-240 (Shimadzu, Kyoto, Japan). P-450 protein concentrations of yeast microsomes were calculated using a millimolar extinction coefficient of 91 (Omura and Sato, *J. Biol. Chem.*, 239:2370-2378 (1964)).

d. In vitro Herbicide Metabolism Assays

Yeast microsomes enriched for a discrete soybean P-450 isozyme were assayed for their capacity to metabolize the ten herbicides and one insecticide listed in Table 3. The reaction mixtures contained 10,000 DPM (100-200 ng) radiolabeled substrate, 0.75 mM NAPDH, 2.5 mg/ml microsomal protein. Total reaction volumes were adjusted to 150 µl with 50 mM phosphate buffer (pH 7.1). The mixtures were incubated under light for 45 minutes at 27°C, arrested with 50 µl acetone and centrifuged at 14 000xg for 2 minutes. Fifty microliters of the supernatants containing radiolabeled alachlor, metolachlor, metribuzin, prosulfuron, chlortoluron, diuron, fluometuron, linuron, or diazinon were spotted onto 250 micron Whatman K6F silica plates. Radiolabeled bentazon and imazaquin-containing samples were spotted onto 200 micron Whatman LKC18F silica gel reversed-phase plates. All plates were developed in a benzene/acetone

2:1 (v/v) solvent system with the exception of prosulfuron, developed in toluene/acetone/acetic acid, 75:20:5 (v/v/v), and bentazon and imazaquin, developed in methanol/75 mM sodium acetate 40:60 (v/v). The developed plates were scanned with a Bioscan System 400 imaging scanner (Bioscan, Washington, DC), and the production of metabolites was determined based on the chromatographic profiles. For microsomes containing the expressed CYP71A10 enzyme, control experiments were also conducted to measure the NADPH-dependency, and the inhibitory effects of CO. CO treatment of the sample was achieved by gentle bubbling of the gas through the reaction mixture for 2 minutes immediately before the assay was initiated by the addition of NADPH.

e. Enzyme Kinetics

Substrate conversion was quantified by a combination of TLC analysis and scintillation spectrometry. The location of the metabolic products on the TLC plates was identified using an imaging scanner, the bands were scraped and analyzed by scintillation spectrometry. The amount of metabolite produced was calculated based on specific activity and scintillation counts. Each assay was repeated at least twice. K_m and V_{max} values were estimated using nonlinear regression analysis.

20

25

30

15

5

10

f. Mass Spectral Analysis

The reaction components used in the *in vitro* fluometuron and linuron metabolism assays were scaled up 50-fold, and the reactions were allowed to proceed for 3 hours. The substrates and the metabolites were extracted 3 times with 20 ml ethyl acetate. The extracts were combined, evaporated to dryness, and the resulting pellet was resuspended in 1 ml acetone. The samples were purified twice using preparative TLC and imaging scanning as described above. Finally, the respective bands were scraped, the compounds were eluted with acetone and flash evaporated.

Fractions of interest were analyzed by liquid chromatography/mass

-24-

spectrometry (LC/MS). Mass spectral measurements were made with a Finnigan TSQ 7000 triple quadruple mass spectrometer (QQQ) equipped with an Atmospheric Pressure Ionization (API) interface fitted with a pneumatically assisted electrospray head (Finnigan MAT, Brennan, Germany). The spray nozzle was operated at 5 kV in the positive ion mode and 4 kV in the negative ion mode. For sample introduction, the TSQ 7000 was equipped with a HPLC solvent delivery system (Perkin-Elmer 410 LC pump), a UV detector (Perkin-Elmer), a stream splitter set at 6:1 with the majority of the effluent flowing to a radioisotope flow monitor (IN/US β-RAM) and the other stream attached to the API interface. Samples were chromatographed on a reverse phase HPLC column (Inertsil 5 ODS2, 150 x 2 mm i.d.). The column was eluted at 0.4 ml/min with 95:5 of 0.1% trifluoroacetic acid in water and 0.1% trifluoroacetic acid in methanol, respectively. Collision induced dissociation experiments (MS/MS) were conducted using argon gas with collision energy in the range of 17.5-30 eV at cell pressures of approximately 0.28 Pa. Signals were captured using a Finnigan 7000 data system.

g. NMR Analysis

5

10

15

20

30

Proton NMR measurements were made on a Bruker AMX-400 NMR spectrometer equipped with either a QNP or inverse probe set at 400.13 MHZ. Spectra were acquired at ambient temperature in acetonitrile- d_3 . Chemical shifts were expressed as parts per million, relative to the resonance of residual acetonitrile protons at 1.93 ppm (δ).

25 h. Tobacco Transformation

A plant expression vector capable of mediating the constitutive expression of CYP71A10 was produced. The GUS open reading frame of the binary expression vector pBI121 (Clontech, Polo Alto, CA) was excised and replaced with the full length CYP71A10 reading frame. This placed the soybean gene under the transcriptional control of the strong constitutive CaMV 35S promoter.

The resulting construct was used to transform Agrobacterium tumefaciens strain LBA 4404 (Holsters et al., *Mol. Gen. Genetics*, 163:181-187 (1988)). Excised leaf discs of Nicotiana tabacum cv SR1 were transformed using the Agrobacterium, and kanamycin-resistant plants were selected as described by Horsch et al. *Science*, 227:1229-1231 (1985). Primary transformants were potted in a standard soil mixture, transferred to a greenhouse and their seed harvested upon maturation.

i. In vivo Herbicide Metabolism Assays

Seeds from primary transgenic tobacco plants transformed with CYP71A10 and control plants transformed with the pBI121 vector were grown in Petri dishes containing MS salts and 100 µg/ml kanamycin. At five weeks post-seeding, kanamycin-resistant plantlets were transplanted into pots containing soil and grown an additional two weeks. Single leaves of approximately 10 cm² in size were excised and their petioles inserted into 100 µl of H₂O containing radiolabeled herbicide. The leaves were placed in a growth chamber maintaining a temperature of 27°C and incubated until the entire volume of the herbicide solution was drawn up by the transpirational stream of the leaves (about 3 hrs). The leaves were subsequently transferred into an Eppendorf tube containing distilled water and further incubated for a total of 14 hours.

[14C]-labeled herbicide was extracted from the leaves by grinding for 5 minutes in 250 μl methanol with a plastic pellet pestle driven by an electric drill. After centrifugation for 3 minutes at 14,000 g, 75 μl of the supernatant was spotted on a Whatman K6F silica plate and developed in a solvent system containing chloroform/ethanol/acetic acid 135:10:15 (v/v/v). The separated herbicide derivatives were visualized using an imaging scanner. Substrate conversion was quantified based on the amount of herbicide absorbed, and the ratios of the parent compound and the produced metabolites determined from the TLC profiles.

30

25

10

15

10

15

20

25

30

i. Herbicide Tolerance

T₁ generation seeds from CYP71A10-transformed tobacco and pBI121-transformed control plants were placed onto Petri dishes containing MS salts and linuron (using its commercial formulation LOROX 50 DF) at active ingredient concentrations ranging from 0.25 to 3.0 μM. Chlortoluron was added at 0, 1.0, 5.0 and 10.0 μM concentrations using a 99.5% pure analytical standard. The Petri dishes were incubated in a growth chamber maintaining a constant temperature of 27°C and a 16/8 hour light/dark cycle. The phytotoxic effects of the treatments were determined visually by comparison to control plants and plants grown in the absence of the herbicide. All treatments were repeated at least twice.

EXAMPLE 2

Isolation of P-450 cDNAs

To isolate cDNAs encoding P-450s from soybean, the PCR strategy described by Meijer et al. (1993) was adapted, using a soybean leaf cDNA library as the template. Degenerate, inosine-containing PCR primers were constructed corresponding to the first nine codons encoding the conserved sequence FLPFGxGxRxCxG (x = any amino acid) (SEQ ID NO:20), which represents an extension of the highly conserved FxxGxxxCxG motif (Bozak et al., *Proc. Natl. Acad. Sci. USA* 87:3904-3908 (1990)) (SEQ ID NO:21). Located near the C-terminal end of the protein, this motif defines the hemebinding region of the protein and may be regarded as a "signature" for P-450 proteins. A second nonspecific primer complementary to the poly(A) tail of the cDNA clones was used in conjunction with these degenerate primers in a PCR amplification assay. PCR amplification products were cloned into a plasmid vector and analyzed by DNA sequencing. Of 86 randomly selected individuals that were sequenced, 15 clones representing 10 unique cDNAs were identified that possessed the conserved cysteine and glycine residues of the signature

consensus (xCxG) (SEQ ID NO:22) immediately following the sequence defined by the degenerate PCR primers. Furthermore, homology searches of the major DNA and protein data bases revealed additional sequence identities to previously reported P-450 sequences for each of the ten unique soybean sequences (data not shown). Because this strategy only allows the recovery of sequence corresponding to the C-terminal portion of the proteins, additional PCR-based techniques were utilized to obtain cDNAs possessing the entire reading frames for each clone. Full length cDNAs were isolated for eight of the 10 individual clones and a near full length cDNA was isolated for an additional clone.

The eight full length and one near full length soybean P-450 cDNAs isolated are described in Table 1. The nomenclature for P-450 genes is based on amino acid sequence identity. Typically, sequences sharing >40% identity are placed in the same family, >55% identity defines members of the same subfamily, and sequences that display >97% identity are assumed to represent allelic variants, although exceptions to these designations have been noted (Nelson et al., *Pharmacogenetics*, 6:1-41 (1996)). According to this system of nomenclature, all of the nine soybean cDNAs were able to be placed within existing P-450 gene families; however, three of the sequences (CYP82C1, CYP83D1 and CYP93C1) defined new subfamilies. Although an increasing number of P-450 gene products have been assigned specific enzymatic functions (reviewed in Schuler, 1996), none of the soybean cDNAs listed in Table 1 could be placed into families for which an *in vivo* function had been determined for any of its members.

In addition to the conserved heme-binding domain described previously, all of the predicted soybean polypeptides possess slight variations of the conserved sequence PEEFxPERF (SEQ ID NO:23) located approximately 30 amino acids forward of the heme-binding motif (Hallahan et al., *Biochem. Soc. Trans.* 21:1068-1073 (1993)). Also characteristic of microsomal P-450s is the presence of an N-terminal noncleavable signal sequence that serves as the membrane anchor. Immediately following this signal-anchor segment in most

5

10

15

20

25

10

microsomal P-450s is a proline-rich region that is believed to form a hinge between the catalytic cytoplasmic domain and the hydrophobic membrane anchor (Halkier, *Phytochemistry* 43:1-21 (1996)). All of the present clones (except CYP97B2) encode proteins pessessing predicted signal sequences; all individuals (except CYP97B2 and CYP82C1) contain readily identifiable proline-rich domains following the signal sequence (Table 1). It is the identification of both of these N-terminal motifs in the CYP83D1 encoded protein (but no Met codon) that indicates that this clone is nearly full length. Interestingly, instead of possessing a predicted signal sequence and proline-rich region, the N-terminus of the polypeptide encoded by clone CYP97B2 contains a motif characteristic of a chloroplast transit peptide (data not shown).

Table 1
Soybean P-450s Isolated Using Degenerate PCR Primers

Name	GenBank Accession #	Length (amino acids)	Closest Match	Identity* %	Membrane Anchor	Proline -rich Region
CYP71A10 (SEQ ID NO:1)	AF022157	513	CYP71A1	51.7	+	+
CYP71D10 (SEQ ID NO:3)	AF022459	510	CYP71D9	50.9	+	+
CYP77A3 (SEQ ID NO:5)	AF022464	513	CYP77A1	69.8	+	+
CYP78A3 (SEQ ID NO:7)	AF022463	523	CYP78A2	53.1	+	+
CYP82C1 (SEQ ID NO:9)	AF022461	532	CYP82A3	51.1	+	_
CYP83D1** (SEQ ID NO:11)	AF022460	516	CYP71A1**	45.7	+	+
CYP93C1 (SEQ ID NO:13)	AF022462	521	CYP93B1	44.5	+	+
CYP97B2 (SEQ ID NO:15)	AF022457	576	CYP97B1	80.8	_	_
CYP98A2 (SEQ ID NO:17)	AF022458	509	CYP98A1	69.7	+	+

¹⁵

^{*}Percent identity between the predicted amino acids sequences of the given soybean P-450 cDNA and the closest match identified from a BLAST search against the major gene and protein databases

^{**} Although this sequence shows a best match to CYP71A1, it matches poorly to some sequences of the CYP71B subfamily. As a result, the tree cluster program places it into the CYP83 family.

EXAMPLE 3

Expression of Soybean P-450 cDNAs in Yeast

Because superfluous 5' untranslated sequences from foreign genes have been shown to be capable of impeding gene expression in yeast (Pompon, 1988), an additional PCR reaction was performed on each clone that enabled the cloning of full length P-450 open reading frames (ORFs) into the yeast expression vector pYeDP60 (V-60) without including any of the endogenous 5' nontranslated flanking sequence (see Methods). For the near full length clone CYP83D1, the 5' primer was also designed to generate an "artificial" Met start codon and a Val second codon at the 5' end of the ORF. Expression in yeast of genes cloned into the V-60 vector is mediated by the strong, galactose-inducible GAL10-CYC1 promoter (Pompon et al., 1995).

Previous studies have revealed that the heterologous expression of P-450 cDNAs in yeast can be greatly enhanced in strains that have been engineered to overexpress endogenous NADPH-dependent cytochrome P-450 reductase (Pompon et al., 1995). In strain W(R), this was accomplished by exchanging the relatively weak endogenous cytochrome P-450 reductase promoter with the same GAL10-CYC1 promoter used in vector V-60 (Truan et al., *Gene* 125:49-55 (1993)). To maximize the heterologous expression of the soybean P-450 cDNAs in yeast, each of the constructs cloned into the V-60 vector was transformed into strain W(R) and microsomes were isolated from cultures that had been induced by galactose.

Reduced-CO difference spectroscopy provides a method to measure the effectiveness of expression of heterologous P-450s in yeast. Microsomal preparations corresponding to five of the soybean constructs (CYP71A10, CYP71D10, CYP77A3, CYP83D1 and CYP98A2) showed characteristic P-450 CO difference spectra with Soret peaks at 450 nm; the profile corresponding to CYP71A10 is shown in Figure 1. No such peaks were observed for the remaining four clones. The specific P-450 content of the five positive

5

10

15

20

25

microsomal preparations varied significantly, ranging from 11 pmol P-450/mg protein for construct CYP83D1 to 252 pmol P-450/mg for clone CYP77A3 as shown in **Table 2**.

Table 2
P-450 Content of Microsomes Isolated from Yeast Overexpressing Various
Soybean CYPs

Clone	CYP content (pmol mg ⁻¹ protein)	
CYP71A10	44	
CYP71D10	15	
CYP77A3	252	
CYP83D1	11	
CYP98A2	13	

10

15

20

25

5

EXAMPLE 4

In vitro Herbicide Assays

To determine whether any of the present soybean P-450 proteins synthesized in yeast displayed significant herbicide metabolic activity, microsomal preparations possessing each of the five soybean P-450s that were effectively expressed in yeast (as judged by their reduced CO difference spectra, see above) were incubated individually with NADPH and radioisotopes of the compounds listed in **Table 3**. These substrates represent six different classes of herbicides and one organophosphate insecticide (diazinon). Upon termination of the reaction, each sample was analyzed by thin layer chromatography (TLC) to reveal potential metabolic breakdown products.

The P-450 proteins expressed from clones CYP71D10, CYP77A3, CYP83D1, and CYP98A2 displayed no apparent *in vitro* metabolic activity against any of the 11 compounds tested (data not shown). In contrast, the P-450 enzyme produced from construct CYP71A10 demonstrated considerable activity

against the phenylurea class of herbicides, but no activity against the remaining compounds. As shown in Figure 2, fluometuron and diuron were converted to a single metabolite; linuron and chlortoluron were transformed into two (a major and a minor) metabolites. Figure 3 shows the chemical structures of the four phenylurea herbicides tested in this study, and the derivatives that have previously been characterized as the first metabolites produced during the detoxification of the respective herbicides in plants known to metabolize these compounds (Voss and Geissbühler, *Proc. Brit. Weed Contr. Conf.* 8:266-268 (1966); Suzuki and Casida, *J. Agric. Food Chem.* 29:1027 (1981); Ryan et al., *Pestic. Biochem. Physiol.* 16:213-221 (1981)).

To further confirm that the herbicide metabolism measured from microsomes of yeast expressing CYP71A10 was attributable to a P-450 activity, additional assays utilizing linuron as the substrate were conducted. As shown in Figure 4, linuron metabolizing activity is reduced approximately 37% in the presence of CO, and no metabolites are observed when NADPH is omitted from the reaction. Activity is also completely abolished upon addition of tetcyclasis, a potent P-450 inhibitor (data not shown). Furthermore, no activity is detected when microsomal preparations are used from yeast cells expressing only the V-60 control plasmid. These results verify that the observed herbicide metabolizing activity is derived from the soybean CYP71A10 cDNA.

The kinetic properties and catalytic activities of the soybean CYP71A10 protein enzyme differed significantly among the four phenylurea-type herbicide substrates. As shown in **Table 4**, turnover rates for fluometuron and linuron were considerably greater than those observed for chlortoluron and diuron. The observed reduced activity for the later two substrates is apparently not the result of decreased binding affinities since the apparent K_m s for chlortoluron and diuron are lower than those measured for fluometuron and linuron.

Table 3

5

10

15

20

25

30 Compounds Used in Metabolism Assays

Common Name	Chemical Family
Alachlor	Acetanilide
Metolachlor	Acetanilide
Bentazon	Benzothiadiazole
Imazaquin	Imidazolinone
Chlortoluron	Pheny!urea
Diuron	Phenylurea
Fluometuron	Phenylurea
Linuron	Phenylurea
Prosulfuron	Sulfonylurea
Metribuzin	as-Triazine
Diazinon	Organophosphate

Table 4
In Vitro Kinetic Parameters of the CYP71A10 Enzyme for Four Phenylurea Substrates

	K _{m, app}	V _{max}	Turnover (min ⁻¹)	
Substrate	(μ M)	(pmol min ⁻¹ mg ⁻¹		
		protein)		
Fluometuron	14.9 (1.0)*	303.6 (10.8)	6.8 (0.24)	
Linuron	9.8 (2.1)	125.6 (12.0)	2.8 (0.27)	
Chlortoluron	1.0 (0.2)	29.4 (2.2)	0.7 (0.05)	
Diuron	1.5 (0.3)	16.8 (1.6)	0.4 (0.04)	

- 5 * Values in parentheses represent standard error.
 - Assays were repeated three times for linuron and twice for all other substrates.
 - Concentration ranges (μM) used were 3.2-27.7 for fluometuron, 3.8-28.3 for linuron, 0.7-4.0 for chlortoluron, and 0.7-3.7 for diuron.

15

20

25

EXAMPLE 5

Analysis of Metabolites

As shown in Figure 2, CYP71A10-mediated degradation of phenylurea herbicides resulted in the accumulation of either one or two metabolites, depending on the particular substrate used. To determine the structure of the metabolites, the single metabolite observed in the fluometuron assay and both the major and minor metabolites generated in the linuron assay were analyzed by liquid chromatography/mass spectroscopy (LC/MS) analysis (results not shown). Analysis of the fluometuron metabolite by LC/MS in positive ion mode resulted in pseudomolecular ions at m/z 219 [(M+H)+, C₉H₉F₃N₂O] and m/z 241 (M+Na)+ that corresponds to a sodium adduct. Daughter ion spectra of m/z 219 produced a prominent m/z 162 fragment ion due to formation of the protonated trifluoromethylaniline $(C_7H_6F_3N+H)^+$. Analysis of the fluometuron metabolite by proton NMR showed a singlet at $\delta 2.71$ which integrated for 3 protons (data not shown). The NMR spectra aromatic resonances were similar to aromatic resonances observed in the parent molecule. Spectra of the fluometuron metabolite were consistent for loss of a methyl group from the parent compound.

-34-

The major linuron metabolite analyzed by LC/MS in the negative ion mode showed a pseudomolecular ion at m/z 233 (M-H) and m/z 235 [(M+2)-H] consistent for a molecule containing two chlorine atoms. Daughter ion spectrum at m/z 233 showed a prominent fragment ion at m/z 160 ($C_6H_4Cl_2N-H$). The major linuron metabolite was 15 mass units less than parent compound which is consistent with loss of a methyl group. The position of methyl loss could not be determined based on mass spectral data alone.

The minor linuron metabolite analyzed by LC/MS gave a pseudomolecular ion at m/z 217 (M-H) and m/z 219 [(M+2)-H] which is consistent for a molecule containing two chlorine atoms. The daughter ion spectrum at m/z 217 showed a prominent fragment ion at m/z 160 which corresponds to formation of the dichloroaniline. The mass spectral data is consistent for the minor linuron metabolite representing N-demethoxy linuron.

These results suggest that the CYP71A10 enzyme expressed in yeast produces the same fluometuron and linuron metabolites as depicted in Figure 3, which shows the first metabolites produced during the detoxification of the respective herbicides in plants that are known to degrade these compounds. The metabolites of chlortoluron and diuron have not been analyzed directly, but the R_f values of the peaks observed during TLC separation are consistent with these species also representing the compounds shown in Figure 3 (ring-hydroxymethyl chlortoluron, N-demethyl chlortoluron and N-demethyl diuron). These results indicate that the CYP71A10 enzyme functions primarily as an N-demethoxylase with respect to fluometuron, linuron and diuron, with some N-demethoxylase activity also observed with linuron. Using chlortoluron as a substrate, the enzyme apparently functions primarily as a methyl-ring hydroxylase and to a lesser extent as an N-demethylase.

EXAMPLE 6

Herbicide Metabolism in Transgenic Tobacco

To determine whether overexpression of the soybean CYP71A10 cDNA

5

10

15

20

25

5

10

15

20

25

in a higher plant system enhances metabolism of phenylurea herbicides, the GUS gene in the binary vector pBI121 was excised and replaced with the CYP71A10 reading frame. This construct placed the CYP71A10 cDNA under the transcriptional control of the constitutive 35S promoter of Cauliflower Mosaic Virus; kanamycin selection was facilitated via the nptII selectable marker. Agrobacterium-mediated transformation of Nicotiana tabacum cv SR1 leaf discs resulted in the recovery of several dozen independent kanamycin-resistant transformants. The plants were subsequently grown to maturity in a greenhouse and allowed to set seed.

For the herbicide metabolism assays, seeds from one randomly selected transgenic line, designated 25/2, were germinated on kanamycin-containing media to eliminate potential nontransgenic segregants. Of 17 germinated seedlings grown, only one individual was inhibited by kanamycin (data not shown). This result suggests that line 25/2 possesses more than one independently segregating transgene. Individual leaves from the 25/2 progeny were excised and incubated with radiolabeled phenylurea herbicides. As shown in Table 5, leaves of the kanamycin-resistant individuals of line 25/2 metabolized all of the four herbicides tested to a much greater extent than the pBI121-transformed control plants.

The relative migrations of the metabolic products revealed by TLC suggest that the products observed in the *in vivo* excised leaf assay are primarily the same as were generated from the *in vitro* assays using yeast microsomes for fluometuron, linuron and diuron (data not shown). For chlortoluron, additional metabolites were also observed. These likely represent combinations of ringmethyl hydroxylated and mono- and di-demethylated species as had been observed by Shiota et al. *Pestic. Biochem. Physiol.* 54:190-198 (1996), in their analysis of chlortoluron-resistant transgenic tobacco that overexpressed the rat CYP1A1 gene. Differences in the ratios of the observed chlortoluron metabolites were also observed between the CYP71A10-transformed and the control plants.

30 Sixty three percent of the metabolites produced in the control leaves was N-

demethyl chlortoluron; in contrast, ring-methyl hydroxy chlortoluron was the most abundant metabolite generated in the CYP71A10-transformed leaves (47%) and only 8% of the metabolites represented N-demethyl chlortoluron.

Table 5

Phenylurea Metabolism after 14 Hours by Excised Leaves of Transgenic

Tobacco Plant 25/2 Progeny

Herbicide ^a	CYP71A10-transformed	Control ^b
	% of herbicide n	netabolized
Fluometuron	91 (4.5)°	15 (0.6)
Linuron	87 (2.0)	12 (2.6)
Chlortoluron	85 (8.1) ^d	39 (7.5) ^d
Diuron	49 (7.0)	20 (2.0)

- (a) Equal amounts of herbicide (1.2 nmol) were added for each experiment.
- (b) Plants transformed with the pBI121 construct were used as controls.
- (c) Values in parentheses represent standard error. A single leaf was assayed from four independent 25/2 plants and three independent control plants.
- (d) The major chlortoluron metabolite in the control plants represented N-demethyl chlortoluron (63%). The metabolites recovered from the CYP71A10-transformed leaves were ring-methyl hydroxy chlortoluron (47%), N-demethyl chlortoluron (8%) and other derivatives (45%).

20

25

10

15

EXAMPLE 7

Herbicide Tolerance

To establish whether enhanced herbicide metabolism leads to an increase in tolerance at the whole plant level, seeds from transgenic plant 25/2 were germinated on an agarose-base medium containing MS salts and varying

concentrations of linuron. Growth of wild-type SR1 plants and transgenic control plants expressing the GUS gene (from vector pBI121) was severely inhibited when exposed to 0.25 µM linuron and completely arrested at concentrations of 0.5 µM and higher (data not shown). As shown in Figure 5, progeny of plant 25/2 grown on media containing no herbicide (Figure 5A) appeared indistinguishable from the same seed grown in the presence of $0.5~\mu M$ linuron (Figure 5C), where only one of 23 germinated seedlings appeared to be inhibited by the herbicide. This ratio appears to be consistent with that observed when seeds from the same parent were grown on selective media containing kanamycin; only one of 17 seedlings failed to grow in the presence of kanamycin. Figure 5B shows control tobacco plants (transformed with vector pBI121), grown on media containing 0.5µM linuron. 25/2 plants tolerant to linuron levels as high as 2.5 µM linuron were observed, although an increasing percentage of the plants showed growth inhibition as the herbicide concentration was increased (Figure 5D). Segregation of the transgene(s) may be leading to variability in expression levels among the progeny of 25/2.

To examine whether the acquisition of herbicide tolerance is unique to line 25/2, seeds from 20 other independent CYP71A10-expressing transgenic plants were similarly germinated and grown on media containing $0.5~\mu \dot{M}$ linuron. Of these, 19 lines gave rise to progeny that were linuron tolerant. The percentage of tolerant individuals for each line varied from approximately 20% to 100% (data not shown). This variation likely represents differences in the copy number, expression levels and segregation of the transgene among the independent lines.

Chlortoluron-tolerance of line 25/2 was also evident. At 1.0 µM herbicide concentration chlortoluron completely arrested the growth of the control plants (Figure 5E). Although growth of the 25/2 plants was modestly inhibited at this herbicide concentration, with the exception of two presumably nontransgenic segregants, the CYP71A10-transformed plants appeared healthy (Figure 5F). In contrast to linuron and chlortoluron, little tolerance of line 25/2

5

10

15

20

25

30

-38-

to fluometuron or diuron was observed. Herbicide concentrations that were injurious to the control plants also inhibited the growth of line 25/2 individuals. Enhanced fluometuron or diuron tolerance was only observed at the very lowest herbicide concentrations necessary to impose growth inhibition in the control plants (data not shown).

5

-39-SEOUENCE LISTING

- (1) GENERAL INFORMATION:
 - (i) APPLICANT: Siminszky, Balazs
 Dewey, Ralph E.
 Corbin, Frederick T.
 - (ii) TITLE OF INVENTION: Novel Cytochrome P-450 Constructs and Methods of Producing Herbicide-Resistant Transgenic Plants
 - (iii) NUMBER OF SEQUENCES: 23
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Virginia C. Bennett
 - (B) STREET: PO Box 37428
 - (C) CITY: Raleigh
 - (D) STATE: North Carolina
 - (E) COUNTRY: USA
 - (F) ZIP: 27627
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
 - (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Bennett, Virginia C.
 - (B) REGISTRATION NUMBER: 37,092
 - (C) REFERENCE/DOCKET NUMBER: 5051-409
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 919-854-1400
 - (B) TELEFAX: 919-854-1401
- (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1838 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 4..1542

-40-

(xi)	SEQUENCE	DESCRIPTION:	SEQ	ID	NO:1:	
------	----------	--------------	-----	----	-------	--

	,,		-														
AAA	ATG Met 1	GCT Ala	CTA Leu	CTA Leu	TCA Ser 5	TCA Ser	GTC Val	CTA Leu	AAG Lys	CAA Gln 10	TTG Leu	CCG Pro	CAT His	GAG Glu	CTA Leu 15	4	8
AGT Ser	TCA Şer	ACC Thr	CAT His	TAC Tyr 20	CTA Leu	ACA Thr	GTT Val	TTC Phe	TTC Phe 25	TGC Cys	ATC Ile	TTC Phe	CTT Leu	ATA ile 30	CTT Leu	9	6
CTT Leu	CAG Gln	CTA Leu	ATA Ile 35	AGA Arg	AGA Arg	AAC Asn	AAA Lys	TAC Tyr 40	AAT Asr.	CTG Leu	CCA Pro	CCA Pro	TCC Ser 45	CCA Pro	CCA Pro	14	4
AAG Lys	ATA Ile	CCC Pro 50	ATA Ile	ATC Ile	GGC Gly	AAT Asn	CTT Leu 55	CAC His	CAG Gln	CTA Leu	GGC Gly	ACA Thr 60	CTG Leu	CCA Pro	CAC His	19	92
CGC Arg	TCC Ser 65	TTT Phe	CAT His	GCA Ala	CTC Leu	TCA Ser	CAC His	AAA Lys	TAT Tyr	GGC Gly	CCT Pro 75	CTC Leu	ATG Met	ATG Met	TTG Leu	. 24	10
CAA Gln 80	TTG Leu	GGT Gly	CAA Gln	ATT Ile	CCA Pro 85	ACC Thr	CTA Leu	GTG Val	GTC Val	TCA Ser 90	TCA Ser	GCT Ala	GAC Asp	GTG Val	GCC Ala 95	28	88
AGA Arg	GAA Glu	ATA Ile	ATC Ile	AAA Lys 100	ACG Thr	CAT ·His	GAT Asp	GTT Val	GTT Val 105	TTC Phe	TCC Ser	AAC Asn	CGC Arg	CGA Arg 110	CAA Gln	33	36
CCT Pro	ACA Thr	GCT Ala	GCT Ala 115	AAA Lys	ATC Ile	TTT Phe	GGT Gly	TAT Tyr 120	GGA Gly	TGC Cys	AAA Lys	GAT Asp	GTG Val 125	GCT Ala	TTC Phe	38	34
GTG Val	TAC Tyr	TAC Tyr 130	CGC Arg	GAA Glu	GAG Glu	TGG Trp	AGA Arg 135	CAA Gln	AAG Lys	ATA Ile	AAG Lys	ACA Thr 140	TGT Cys	AAG Lys	GTT Val	43	32
GAG Glu	CTT Leu 145	Met	AGT Ser	CTG Leu	AAG Lys	AAG Lys 150	GTG Val	CGG Arg	TTG Leu	TTT Phe	CAT His 155	TCC Ser	ATT Ile	AGA Arg	CAA Gln	4.8	30
GAA Glu 160	Val	GTT Val	ACA Thr	GAG Glu	TTG Leu 165	Val	GAA Glu	GCT Ala	ATA Ile	GGT Gly 170	Glu	GCG Ala	TGT Cys	GGT Gly	AGT Ser 175	52	28
GAA Glu	AGA Arg	CCA Pro	TGT Cys	GTG Val 180	AAT Asn	CTG Leu	ACT Thr	GAG Glu	ATG Met 185	CTG Leu	ATG Met	GCA Ala	GCA Ala	TCG Ser 190	Asn	5	76
				Arg					Arg	AAG Lys				Ala	TGT Cys	62	24
GGT Gly	GGT Gly	AGT Ser 210	Gly	AGT Ser	AGC Ser	AGC Ser	TTT Phe 215	Ala	GCG Ala	TTG Leu	GGA Gly	AGA Arg 220	Lys	ATT	ATG Met	6	72
AGA	. CTA	TTA	TCG	GCT	TTC	AGC	GTG	GGT	GAT	TTC	TTC	CCT	TCG	TTG	GGT	7:	20

-41-

									-41							,
Arg	Leu 225	Leu	Ser	Ala	Phe	Ser 230	Val	Gly	Asp	Phe	Phe 235	Pro	Ser	Leu	Gly	
TGG Trp 240	GTT Val	GAC Asp	TAT Tyr	CTG Leu	ACT Thr 245	GGC Gly	TTA Leu	ATT Ile	CCA Pro	GAG Glu 250	ATG Met	AAA Lys	ACC Thr	ACG Thr	TTT Phe 255	768
CTC Leu	GCA Ala	GTA Val	GAT Asp	GCT Ala 260	TTC Phe	CTT Leu	GAT Asp	GAG Glu	GTA Val. 265	ATT Ile	GCA Ala	GAA Gl.u	CAC His	GAG Glu 270	AGC Ser	816
AGT Ser	AAC Asn	AAG Lys	AAG Lys 275	AAT Asn	GAT Asp	GAC Asp	TTC Phe	TTG Leu 280	GGG Gly	ATA Ile	CTT Leu	CTT Leu	CAA Gln 285	CTT Leu	CAA Gln	854
GAA Glu	TGT Cys	GGG Gly 290	AGG Arg	CTT Leu	GAC Asp	TTT Phe	CAG Gln 295	CTC Leu	GAC Asp	CGA Arg	GAT Asp	AAC Asn 300	CTC Leu	AAA Lys	GCA Ala	912
ATC Ile	CTA Leu 305	GTG Val	GAC Asp	ATG Met	ATA Ile	ATA Ile 310	GGT Gly	GGG Gly	AGT Ser	GAC Asp	ACT Thr 315	ACT Thr	TCA Ser	ACA Thr	ACT Thr	960
CTA Leu 320	GAA Glu	TGG Trp	ACT Thr	TTT Phe	GCG Ala 325	GAG Glu	TTC Phe	CTT Leu	AGA Arg	AAT Asn 330	CCA Pro	AAT Asn	ACC Thr	ATG Met	AAG Lys 335	1008
AAA Lys	GCT Ala	CAA Gln	GAA Glu	GAG Glu 340	GTA Val	AGA Arg	AGA Arg	GTG Val	GTG Val 345	GGA Gly	ATC Ile	AAT Asn	TCC Ser	AAA Lys 350	GCA Ala	1056
GTA Val	CTG Leu	GAT Asp	GAA Glu 355	AAT Asn	TGT Cys	GTG Val	AAT Asn	CAA Gln 360	ATG Met	AAC Asn	TAC Tyr	TTG Leu	AAA Lys 365	TGT Cys	GTA Val	1104
GTC Val	AAA Lys	GAA Glu 370	ACT Thr	TTG Leu	AGA Arg	TTA Leu	CAT His 375	CCA Pro	CCC Pro	CTT Leu	CCT Pro	CTT Leu 380	TTG Leu	ATT Ile	GCT Ala	<u>1</u> 152
CGA Arg	GAG Glu 385	ACA Thr	TCA Ser	Ser	Ser	GTA Val 390	Lys	Leu	Arg	Gly	Tyr	Asp	ATT Ile	CCC Pro	GCA Ala	1200
AAA Lys 400	ACA Thr	ATG Met	GTA Val	TTT Phe	ATC Ile 405	AAT Asn	GCA Ala	TGG Trp	GCG Ala	ATC Ile 410	CAG Gln	AGG Arg	GAT Asp	CCT Pro	GAA Glu 415	1248
TTA Leu	TGG Trp	GAT Asp	GAT Asp	CCT Pro 420	Glu	GAA Glu	TTT Phe	ATT Ile	CCC Pro 425	GAA Glu	AGA Arg	TTT Phe	GAA Glu	ACT Thr 430	AGC Ser	1296
CAA Gln	GTT Val	GAT Asp	CTT Leu 435	AAT Asn	GGA Gly	CAA Gln	GAT Asp	TTT Phe 440	Gln	TTA Leu	ATT	CCG Pro	TTC Phe 445	GGT Gly	ATT Ile	1344
GGG Gly	AGA Arg	AGG Arg 450	Gly	TGC Cys	CCT Pro	GCA Ala	ATG Met 455	Ser	TTT Phe	GGA Gly	CTT Leu	GCT Ala 460	TCA Ser	ACT Thr	GAG Glu	1392

PCT/US98/20807 WO 99/19493

									~+2	_						
TAT Tyr	GTT Val 465	CTT Leu	GCT Ala	AAT Asn	CTT Leu	TTG Leu 470	TAT Tyr	TGG Trp	TTC Phe	AAT Asn	TGG Trp 475	AAT Asn	ATG Met	TCC Ser	GAG Glu	1440
TCT Ser 480	GGA Gly	CGT Arg	ATA Ile	TTG Leu	ATG Met 485	CAC His	AAC Asn	ATT Ile	GAC Asp	ATG Met 490	AGT Ser	GAG Glu	ACA Thr	AAT Asn	GGA Gly 495	1488
CTC Leu	ACT Thr	GTC Val	AGT Ser	AAG Lys 500	AAA Lys	GTA Val	CCA Pro	CTT Leu	CAT His 505	CTT Leu	GAA Glu	CCA Pro	GAA Glu	CCA Pro 510	TAT Tyr	1536
	ACA Thr	TGA:	rcat:	TTC A	ACAT'	ratg(CA TO	GTTT	GCA)	A CAC	CCTA!	ГААА	GAG'	TATA	GAT	1592
CTG	GAAG'	rac :	rtca.	LTTT.	AG T	AATG(GATG'	KAA 1	AAGC'	TATA	CAA	raag:	AAG 1	rgct/	AACAAG	1652
CTA	GGAT	ATG A	AGCA!	TTTA'	TG G	AGTA	ACGA(G TGA	AGGT	rcca	AAG	AGTC'	raa '	TTAC'	TCGTCT	1712
CTT	JAAC.	ATT (GTTA:	TATT	TG T	TTTC:	TTGC	A GT	TTGT	TAAT	CTT	rtga.	ATA (GTTG'	TTTCAC	1772
ATT'	TATT'	TTT (GTAT	GGTT	TG T	rggt	ATGT	r GT	GGAA	GCG	TTA	GTAA	AAA '	TTTG'	TGGTGT	1832
стт																1838

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 513 amino acids

 - (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Leu Leu Ser Ser Val Leu Lys Gln Leu Pro His Glu Leu Ser 10

Ser Thr His Tyr Leu Thr Val Phe Phe Cys Ile Phe Leu Ile Leu Leu 25 20

Gln Leu Ile Arg Arg Asn Lys Tyr Asn Leu Pro Pro Ser Pro Pro Lys 40

Ile Pro Ile Ile Gly Asn Leu His Gln Leu Gly Thr Leu Pro His Arg 55

Ser Phe His Ala Leu Ser His Lys Tyr Gly Pro Leu Met Met Leu Gln 75 70

Leu Gly Gln Ile Pro Thr Leu Val Val Ser Ser Ala Asp Val Ala Arg 90 85

Glu Ile Ile Lys Thr His Asp Val Val Phe Ser Asn Arg Arg Gln Pro 105

Thr Ala Ala Lys Ile Phe Gly Tyr Gly Cys Lys Asp Val Ala Phe Val

-43-

										72	_					
			115					120					125			
-	ryr	Tyr 130	Arg	Glu	Glu	Trp	Arg 135	Gln	Lys	Ile	Lys	Thr 140	Cys	Lys	Val	Glu
	Leu 145	Met	Ser	Leu	Lys	Lys 150	Val	Arg	Leu	Phe	His 155	Ser	Ile	Arg	Gln	Glu 160
7	/al	Val	Thr	Glu	Leu 165	Val	Glu	Ala	Ile	Gly 170	Glu	Ala	Cys	Gly	Ser 175	Glu
1	Arg	Pro	Cys	Val 180	Asn	Leu	Thr	Glu	Met 185	Leu	Met	Ala	Ala	Ser 190	Asn	Asp
	Ile	Val	Ser 195	Arg	Cys	Val	Leu	Gly 200	Arg	Lys	Cys	Asp	Asp 205	Ala	Cys	Gly
	Gly	Ser 210	Gly	Ser	Ser	Ser	Phe 215	Ala	Ala	Leu	Gly	Arg 220	Lys	Ile	Met	Arg
	Leu 225	Leu	Ser	Ala	Phe	Ser 230	Val	Gly	Asp	Phe	Phe 235	Pro	Ser	Leu	Gly	Trp 240
	Val	Asp	Tyr	Leu	Thr 245	Gly	Leu	Ile	Pro	Glu 250	Met	Lys	Thr	Thr	Phe 255	Leu
	Ala	Val	Asp	Ala 260	Phe	Leu	Asp	Glu	Val 265	Ile	Ala	Glu	His	Glu 270	Ser	Ser
	Asn	Lys	Lys 275	Asn	Asp	Asp	Phe	Leu 280	Gly	Ile	Leu	Leu	Gln 285	Leu	Gln	Glu
	Cys	Gly 290		Leu	Asp	Phe	Gln 295	Leu	Asp	Arg	Asp	Asn 300	Leu	Lys	Ala	Ile
	Leu 305	Val	Asp	Met	Ile	Ile 310		Gly	Ser	Asp	Thr 315	Thr	Ser	Thr	Thr	Leu 320
	Glu	Trp	Thr	Phe	Ala 325	Glu	Phe	Leu	Arg	Asn 330	Pro	Asn	Thr	Met	Lys 335	Lys
	Ala	Gln	Glu	Glu 340		Arg	Arg	Val	Val 345	Gly	·Ile	Asn	Ser	Lys 350	Ala	Val
	Leu	Asp	Glu 355		Cys	Val	Asn	Gln 360		Asn	Tyr	Leu	Lys 365	Cys	Val	Val
	Lys	Glu 370		Leu	Arg	Leu	His 375		Pro	Leu	Pro	Leu 380		lle	: Ala	Arg
	Glu 385		Ser	Ser	Ser	Val 390		Leu	Arg	Gly	7 Tyr 395	Asp	Ile	Pro	Ala	400
	Thr	Met	: Val	Phe	1le 405		Ala	Trp	Ala	11e		a Arg	Asp	Prc	415	Leu ;
	Trp	Asp	Asp	Pro 420		Glu	Phe	: Ile	Pro 425		ı Arg	g Phe	e Glu	430	Ser	Glr
	Val	. Asp	Let	ı Asn	Gly	Gln	Asp	Phe	Glr	. Lei	ı Ile	e Pro	Phe	e Gly	/ Ile	e Gly

-44-

435 440 445

Arg Arg Gly Cys Pro Ala Met Ser Phe Gly Leu Ala Ser Thr Glu Tyr 450 455 460

Val Leu Ala Asn Leu Leu Tyr Trp Phe Asn Trp Asn Met Ser Glu Ser 465 470 475 480

Gly Arg Ile Leu Met His Asn Ile Asp Met Ser Glu Thr Asn Gly Leu 485 490 495

Thr Val Ser Lys Lys Val Pro Leu His Leu Glu Pro Glu Pro Tyr Lys 500 505 510

Thr

- (2) INFORMATION FOR SEQ ID NO:3:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1691 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 16..1545
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCTAGATCTA TCATC							TCT Ser	51
	1		5			10		

- TTA GTT CAA AGA TCG GAT TCC AAA ACC TCC TCT ACC TGC AAA TTG CCC
 Leu Val Gln Arg Ser Asp Ser Lys Thr Ser Ser Thr Cys Lys Leu Pro
 30 40
- CCA GGA CCA AGG ACA CTA CCT CTC ATA GGG AAC ATA CAC CAG ATT GTT

 Pro Gly Pro Arg Thr Leu Pro Leu Ile Gly Asn Ile His Gln Ile Val

 45 50 55 60
- GGC TCA CTG CCG GTT CAT TAC TAC TTA AAA AAT TTG GCA GAT AAG TAT

 Gly Ser Leu Pro Val His Tyr Tyr Leu Lys Asn Leu Ala Asp Lys Tyr

 70

 75
- GGT CCA TTA ATG CAT CTA AAA CTA GGA GAG GTG TCC AAC ATC ATA GTC

 Gly Pro Leu Met His Leu Lys Leu Gly Glu Val Ser Asn Ile Ile Val

 80 85 90
- ACT TCC CCA GAA ATG GCC CAA GAG ATT ATG AAG ACA CAT GAT CTC AAC 339

-45-

							-45									
	Asn	Leu	Asp	His 105	Thr	Lys	Met	Ile	Glu 100	Gln	Ala	Met	Glu	Pro 95	Ser	Thr
387	AAC Asn	TAC Tyr	TCT Ser	GTT Val	ATA Ile 120	AGA Arg	TCT Ser	TTG Leu	GTA Val	TTT Phe 115	GAC Asp	CCA Pro	AGG Arg	GAT Asp	TCT Ser 110	TTC Phe
435	CTA Leu 140	CAA Gln	AGG Arg	TGJ Trp	TAT Tyr	GAC Asp 135	GGA Gly	CAT His	CAA Gln	AGT Ser	TTC Phe 130	GTC Val	ATT Ile	GGC Gly	TCT Ser	GGT Gly 125
483	TCT Ser	CAG Gln 155	GTG Val	CGC Arg	AAG Lys	GCA Ala	ACA Thr 150	CTA Leu	TTA Leu	GAG Glu	GTA Val	ACA Thr 145	TGC Cys	ATA Tle	AAG Lys	AGA Arg
531	ATA Ile	AAA Lys	AAA Lys 170	GTT Val	CTA Leu	GAA Glu	GCA Ala	GTG Val 165	GAG Glu	GAG Glu	GAA Glu	AGA Arg	ATA Ile 160	TCC Ser	CGG Arg	TTT Phe
579	CAG Gln	ACC Thr	CTC Leu	AAT Asn 185	TTT Phe	ATT Ile	TCC Ser	GGG Gly	GGG Gly 180	GAA Glu	GAA Glu	AGT Ser	GCA Ala	ACT Thr 175	GCA Ala	GCT Ala
627	GGT Gly	TTT Phe	GCT Ala	GCG Ala	CGA Arg 200	GCA Ala	GCG Ala	ATA Ile	GGG Gly	TTT Phe 195	ACT Thr	ATG Met	TCA Ser	TAC Tyr	ATT Ile 190	AGC Ser
675	CAA Gln 220	AAA Lys	CAT His	ATG Met	AAC Asn	TCA Ser 215	ATA Ile	TTC Phe	GTG Val	CAA Gln	CAA Gln 210	TAC Tyr	AGA Arg	AGC Ser	Lys	AAA Lys 205
723	AGT Ser	TCT Ser 235	CCT Pro	TAT Tyr	CTC Leu	GAT Asp	GCT Ala 230	GTT Val	TCT Ser	TTT Phe	GGG Gly	GGA Gly 225	CTG Leu	CTT Leu	ATG Met	TTG Leu
. 771	CAT His	GTG Val	AAA Lys 250	GAA Glu	CTT Leu	AAA Lys	GGG	ACG Thr 245	GCG Ala	GGG Gly	ATG Met	ATG Met	CAA Gln 240	TTT Phe	GTĞ Val	AGA Arg
819	AAT Asn	AAA Lys	CAC His	GAG Glu 265	GAC Asp	ATC Ile	ATC Ile	GAC Asp	CAA Gln 260	TTG Leu	GTG Val	AGG Arg	GAT Asp	ACA Thr 255	GTG Val	AGA Arg
867	GAT Asp	GTT Val	CTA Leu	GAT Asp	GAA Glu 280	GTG Val	GCA Ala	GAA Glu	CGT Arg	GAG Glu 275	GAG Glu	AGC Ser	AGC Ser	Arg	AAC Asn 270	AGA Arg
915		GAT Asp													Leu	
963	ACA Thr	GAA Glu 315	GGC Gly	GGA Gly	GGT Gly	ATT Ile	TTC Phe 310	ATA Ile	GAC Asp	CAG Gln	ATC Ile	GTC Val 305	GCC Ala	AAA Lys	ATT Ile	AAC Asn
1011	CCG Pro	AAC Asn	AGA Arg 330	ATA Ile	TTG Leu	GAA Glu	TCA Ser	ATG Met 325	GGG Gly	TGG Trp	GAA Glu	GTG Val	GTT Val 320	TCT Ser	TCT	TCA Ser

-46-

									-40	-						
AGG Arg	GTG Val	ATG Met 335	GAA Glu	GAA Glu	GCA Ala	CAA Gln	GCA Ala 340	GAG Glu	GTG Val	AGA Arg	AGA Arg	GTG Val 345	TAT Tyr	GAT Asp	AGC Ser	1059
AAG Lys	GGA Gly 350	TAT	GTG Val	GAT Asp	GAG Glu	ACA Thr 355	GAA Glu	TTG Leu	CAC His	CAA Gln	TTG Leu 360	ATA Ile	TAC Tyr	TTA Leu	AAG Lys	1107
TCC Ser 365	ATC Ile	ATC Ile	AAA Lys	GAA Glu	ACC Thr 370	ATG Met	AGG Arg	TTA Leu	CAT His	CCA Pro 375	CCT Pro	GTG Val	CCA Pro	TTG Leu	TTA Leu 380	1155
GTT Val	CCT Pro	AGA Arg	GTA Val	AGT Ser 385	AGA Arg	GAA Glu	AGG Arg	TGC Cys	CAA Glm 390	ATC Ile	AAT Asn	GGA Gly	TAT Tyr	GAG Glu 395	ATA Ile	1203
CCC Pro	TCT Ser	AAG Lys	ACT Thr 400	AGG Arg	ATC Ile	ATT Ile	ATC Ile	AAT Asn 405	GCT Ala	TGG Trp	GCA Ala	ATT Ile	GGA Gly 410	AGG Arg	AAT Asn	1251
													AGG Arg			1299
													ATC Ile			1347
GGT Gly 445	GCT Ala	GGA Gly	AGG Arg	AGG Arg	ATC Ile 450	TGC Cys	CCC Pro	GGC Gly	ATT Ile	ACA Thr 455	TTT Phe	GCC Ala	ATA Ile	CCC Pro	AAC Asn 460	1395
													TGG Trp			1443
CCC Pro	AAT Asn	AAA Lys	ATG Met 480	AAG Lys	AAT Asn	GAA Glu	GAA Glu	CTT Leu 485	GAC Asp	ATG Met	ACG Thr	GAG Glu	TCA Ser 490	AAT Asn	GGA Gly	1491
ATT Ile	ACT Thr	TTA Leu 495	CGA Arg	AGA Arg	CAA Gln	AAT Asn	GAC Asp 500	CTC Leu	TGC Cys	TTG Leu	ATT	CCC Pro 505	ATT Ile	ACT Thr	CGT Arg	1539
	CCT Pro 510		AATG	TAT (GAAC	AATT.	AA T	GTCA	TAAA	C TA	TTTA	AGTT	TTA	TCTT	TTA	1595
CTA	CTTC	CAG	CATT	TCGT	AA T'	TGGA	CAAT	G AC	TATG	ATTA	ACT	TAAG	TTA	CTTC	CTTATG	1655
ATT	AACT	TGA	CATA	TGAA	TG A	ACAT	TTCT	a ag	ATAA							1691

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 510 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

-47-

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Val Met Glu Leu His Asn His Thr Pro Phe Ser Ile Tyr Phe Ile
5 10 15

Thr Ser Ile Leu Phe Ile Phe Phe Val Phe Phe Lys Leu Val Gln Arg
20 25 30

Ser Asp Ser Lys Thr Ser Ser Thr Cys Lys Leu Pic Pro Gly Pro Arg

Thr Leu Pro Leu Ile Gly Asn Ile His Gln Ile Val Gly Ser Leu Pro 50 55 60

Val His Tyr Tyr Leu Lys Asn Leu Ala Asp Lys Tyr Gly Pro Leu Met 65 70 75 80

His Leu Lys Leu Gly Glu Val Ser Asn Ile Ile Val Thr Ser Pro Glu 85 90 95

Met Ala Gln Glu Ile Met Lys Thr His Asp Leu Asn Phe Ser Asp Arg

Pro Asp Phe Val Leu Ser Arg Ile Val Ser Tyr Asn Gly Ser Gly Ile 115 120 125

Val Phe Ser Gln His Gly Asp Tyr Trp Arg Gln Leu Arg Lys Ile Cys 130 135 140

Thr Val Glu Leu Leu Thr Ala Lys Arg Val Gln Ser Phe Arg Ser Ile 145 150 155 160

Arg Glu Glu Val Ala Glu Leu Val Lys Lys Ile Ala Ala Thr Ala 165 170 175

Ser Glu Glu Gly Gly Ser Ile Phe Asn Leu Thr Gln Ser Ile Tyr Ser 180 185 190

Met Thr Phe Gly Ile Ala Ala Arg Ala Ala Phe Gly Lys Lys Ser Arg

Tyr Gln Gln Val Phe Ile Ser Asn Met His Lys Gln Leu Met Leu Leu 210 215 220

Gly Gly Phe Ser Val Ala Asp Leu Tyr Pro Ser Ser Arg Val Phe Gln
225 230 235 240

Met Met Gly Ala Thr Gly Lys Leu Glu Lys Val His Arg Val Thr Asp 245 250 255

Arg Val Leu Gln Asp Ile Ile Asp Glu His Lys Asn Arg Asn Arg Ser 260 265 270

Ser Glu Glu Arg Glu Ala Val Glu Asp Leu Val Asp Val Leu Leu Lys 275 280 285

Phe Gln Lys Glu Ser Glu Phe Arg Leu Thr Asp Asp Asn Ile Lys Ala

-48-

290 295 300

Val Ile Gln Asp Ile Phe Ile Gly Gly Glu Thr Ser Ser Ser Val 305 310 315 320

Val Glu Trp Gly Met Ser Glu Leu Ile Arg Asn Pro Arg Val Met Glu 325 330 335

Glu Ala Gln Ala Glu Val Arg Arg Val Tyr Asp Ser Lys Gly Tyr Val 340 345 350

Asp Glu Thr Glu Leu His Gln Leu Ile Tyr Leu Lys Ser Ile Ile Lys 355 360 365

Glu Thr Met Arg Leu His Pro Pro Val Pro Leu Leu Val Pro Arg Val 370 375 380

Ser Arg Glu Arg Cys Gln Ile Asn Gly Tyr Glu Ile Pro Ser Lys Thr 385 390 395 400

Arg Ile Ile Ile Asn Ala Trp Ala Ile Gly Arg Asn Pro Lys Tyr Trp 405 410 415

Gly Glu Thr Glu Ser Phe Lys Pro Glu Arg Phe Leu Asn Ser Ser Ile 420 425 430

Asp Phe Arg Gly Thr Asp Phe Glu Phe Ile Pro Phe Gly Ala Gly Arg 435 440 445

Arg Ile Cys Pro Gly Ile Thr Phe Ala Ile Pro Asn Ile Glu Leu Pro 450 455 460

Leu Ala Gln Leu Leu Tyr His Phe Asp Trp Lys Leu Pro Asn Lys Met 465 470 475 480

Lys Asn Glu Glu Leu Asp Met Thr Glu Ser Asn Gly Ile Thr Leu Arg
485 490 495

Arg Gln Asn Asp Leu Cys Leu Ile Pro Ile Thr Arg Leu Pro 500 505 510

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1644 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 4..1542
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAA ATG GCC ACT CTT TCC TCC TAC GAC CAC TTC ATC TTC ACT GCC TTA

-49-

									-49							
	Met 1	Ala	Thr	Leu	Ser 5	Ser	Tyr	Asp	His	Phe 10	Ile	Phe	Thr	Ala	Leu 15	
GCT Ala	TTC Phe	TTC Phe	ATA Ile	TCT Ser 20	GGC Gly	CTA Leu	ATT Ile	TTC Phe	TTC Phe 25	CTC Leu	AAA Lys	CAG Gln	AAA Lys	TCC Ser 30	AAA Lys	96
TCC Ser	AAA Lys	AAG Lys	TIC Phe 35	AAC Asn	CTC Leu	CCT Pro	CCA Pro	GGA Gly 40	CCC Pro	CCC Pro	GGG Gly	TGG Trp	CCT Pro 45	ATT Ile	GTT Val	144
GGG. Gly	AAC Asn	CTC Leu 50	TTC Phe	CAA Gln	GTT Val	GCT Ala	CGT Arg 55	TCT Ser	GGG Gly	AAA Lys	CCT Pro	TTC Phe 60	TTT Phe	GAG Glu	TAT Tyr	192
GTG Val	AAC Asn 65	GAT Asp	GTG Val	AGA Arg	CTC Leu	AAA Lys 70	TAT Tyr	GGC Gly	TCA Ser	ATC Ile	TTC Phe 75	ACC Thr	CTC Leu	AAG Lys	ATG Met	240
GGA Gly 80	ACA Thr	AGG Arg	ACC Thr	ATG Met	ATC Ile 85	ATC Ile	CTC Leu	ACC Thr	GAC Asp	GCA Ala 90	AAA Lys	CTG Leu	GTC Val	CAC His	GAG Glu 95	288
GCC Ala	ATG Met	ATC Ile	CAA Gln	AAG Lys 100	GGT Gly	GCA Ala	ACC Thr	TAC Tyr	GCC Ala 105	ACC Thr	AGG Arg	CCC Pro	CCC Pro	GAG Glu 110	AAC Asn	336
CCC Pro	ACC Thr	AGA Arg	ACC Thr 115	ATC Ile	TTC Phe	AGT Ser	GAA Glu	AAC Asn 120	AAG Lys	TTC Phe	ACC Thr	GTG Val	AAT Asn 125	GCA Ala	GCG Ala	384
ACC Thr	TAT	GGC Gly 130	CCC Pro	GTG Val	TGG Trp	AAG Lys	TCG Ser 135	CTG Leu	AGG Arg	AGG Arg	AAC Asn	ATG Met 140	GTG Val	CAG Gln	AAC Asn	432
ATG Met	CTC Leu 145	Ser	TCA Ser	ACA Thr	AGA Arg	CTT Leu 150	AAG Lys	GAG Glu	TTT Phe	CGC Arg	AGT Ser 155	GTT Val	CGG Arg	GAC Asp	AAT Asn	480
GCG Ala 160	Met	GAC Asp	AAG Lys	CTC Leu	ATC Ile 165	AAC Asn	AGA Arg	CTC Leu	AAG Lys	GAC Asp 170	GAG Glu	GCC Ala	GAG Glu	AAG Lys	AAT Asn 175	528
AAC Asn	GGC Gly	GTG Val	GTT Val	TGG Trp 180	GTG Val	CTC Leu	AAG Lys	GAT Asp	GCC Ala 185	AGG Arg	TTT Phe	GCT Ala	GTT Val	TTT Phe 190	TGC Cys	576
ATA Ile	CTI Leu	GTG Val	GCT Ala 195	Met	TGT Cys	TTT Phe	GGT Gly	CTT Leu 200	Glu	ATG Met	GAT Asp	GAG Glu	GAG Glu 205	Thr	GTG Val	624
GAG Glu	AGA Arg	ATA Ile 210	Asp	CAG Gln	GTT Val	ATG Met	AAG Lys 215	Ser	GTT Val	CTC Leu	ATC Ile	ACT Thr 220	Leu	GAC Asp	CCG Pro	672
AGA Arg	ATT 11e 225	. Asp	GAC Asp	TAT	CTT Leu	CCA Pro 230	Ile	'CTA Leu	AGC Ser	CCC	TTT Phe 235	Phe	TCA Ser	AAG Lys	CAA Gln	720

									-50	_						• •
AGA Arg 240	AAG Lys	AAA Lys	GCC Ala	TTG Leu	GAG Glu 245	GTT Val	CGC Arg	AGA Arg	GAA	CAG	GTT Val	GAG Glu	TTC Phe	TTA Leu	GTT Val 255	768
CCA Pro	ATT Ile	ATA Ile	GAA Glu	CAA Gln 260	AGA Arg	AGA Arg	AGA Arg	GCA Ala	ATT Ile 265	CAA Gln	AAC Asn	CCT Pro	GGG Gly	TCA Ser 270	GAT Asp	816
CAC His	ACC Thr	GCC Ala	ACA Thr 275	ACG Thr	TTT Phe	TCC Ser	TAC Tyı	CTA Leu 280	GAC Asp	ACA Thr	CTT Leu	TTT Phe	GAC Asp 285	CTC Leu	AAA Lys	864
GTT Val	GAA Glu	GGG Gly 290	AAG Lys	AAA Lys	TCA Ser	GCA Ala	CCC Pro 295	TCT Ser	GAT Asp	GCA Ala	GAA Glu	TTG Leu 300	GTG Val	TCT Ser	TTA Leu	912
TGC Cys	TCA Ser 305	GAG Glu	TTT Phe	CTT Leu	AAC Asn	GGT Gly 310	GGC Gly	ACA Thr	GAC Asp	ACA Thr	ACA Thr 315	GCA Ala	ACA Thr	GCG Ala	GTT Val	960
GAG Glu 320	TGG Trp	GGC Gly	ATA Ile	GCA Ala	CAG Gln 325	CTC Leu	ATA Ile	GCG Ala	AAC Asn	CCT Pro 330	AAC Asn	GTT Val	CAG Gln	ACA Thr	AAG Lys 335	1008
CTG Leu	TAC Tyr	GAG Glu	GAA Glu	ATA Ile 340	Lys	AGA Arg	ACG Thr	GTG Val	GGA Gly 345	GAG Glu	AAG Lys	AAG Lys	GTG Val	GAT Asp 350	GAA Glu	1056
AAG Lys	GAC Asp	GTT Val	GAG Glu 355	AAA Lys	ATG Met	CCA Pro	TAC Tyr	CTA Leu 360	CAC His	GCT Ala	GTG Val	GTG Val	AAG Lys 365	GAG Glu	CTT Leu	1104
Leu	Arg	Lys 370	CAC His	Pro	Pro	Thr	His 375	Phe	Val	Leu	Thr	His 380	Ala	Val	Thr	1152
GAG Glu	CCC Pro 385	ACC Thr	ACT Thr	TTG Leu	GGA Gly	GGG Gly 390	TAT	GAC Asp	ATA Ile	CCA Pro	ATT Ile 395	GAT Asp	GCA Ala	AAT Asn	GTT Val	1200
Glu 400	Val	Tyr	ACA Thr	Pro	Ala 405	Ile	Ala	Glu	Asp	Pro 410	Lys	Asn	Trp	Leu	Asn 415	1248
Pro	Glu	Lys	TTT Phe	Asp 420	Pro	Glu	Arg	Phe	Ile 425	Ser	Gly	Gly	Glu	Glu 430	Ala	1296
			GGG Gly 435						Met						GGG Gly	1344
Arg	Arg	Ile 450		Pro	Gly	Leu	Ala 455	Met	Ala	Thr	Val	His 460	Ile	His	Leu	1392
ATG Met	ATG Met 465	Ala	AGG Arg	ATG Met	GTG Val	CAG Gln 470	Glu	TTT	GAG Glu	TGG Trp	GGT Gly 475	Ala	TAC	CCT Pro	CCA Pro	1440

-51-

GAG Glu 480	AAG Lys	AAG Lys	ATG Met	GAT Asp	TTC Phe 485	ACT Thr	GGC Gly	AAG Lys	TGG Trp	GAG Glu 490	TTC Phe	ACT Thr	GTG Val	GTC Val	ATG Met 495	1488
AAG Lys	GAG Glu	TCT Ser	CTA Leu	AGA Arg 500	GCA Ala	ACC Thr	ATC Ile	AAA Lys	CCA Pro 505	AGA Arg	GGA Gly	GGA Gly	GAA Glu	AAA Lys 510	GTG Val	1536
AAG Lys		TAAA	LTTT	TC C	TGCI	TCTA	AT TO	TTC	rgggi	TTT	TAAAT	TTTC	ACAG	GACA)	ACA	1592
TAAA	TAT	TAT T	GCTA	ratt.	C AI	CATO	ATA:	TA T	TATA	CAT	CATO	CATGO	GTT A	4C		1644
(2)	INFO	RMAT	CION	FOR	SEQ	ID N	10:6:	•								,
	•	(i) S	(B)	LEN	IGTH:	ACTE : 513 :minc :SY: 1	ami aci	ino a Id	: acids	3						
			OLEC													
			EQUE													•
Met 1	Ala	Thr	Leu	Ser 5	Ser	Tyr	Asp	His	Phe 10	Ile	Phe	Thr	Ala	Leu 15	Ala	
Phe	Phe	Ile	Ser 20	Gly	Leu	Ile	Phe	Phe 25	Leu	Lys	Gln	Lys	Ser 30	Lys	Ser	
Lys	Lys	Phe 35	Asn	Leu	Pro	Pro	Gly 40	Pro	Pro	Gly	Trp	Pro 45	Ile	Val	Gly	\$\$ ₩.
Asn	Leu 50	Phe	Gln	Val	Ala	Arg 55	Ser	Gly	Lys	Pro	Phe 60	Phe	Glu	Tyr	Val	
Asn 65	Asp	Val	Arg	Leu	Lys 70	Tyr	Gly	Ser	Ile	Phe 75	Thr	Leu	Lys	Met	Gly 80	
Thr	Arg	Thr	Met	Ile 85	Ile	Leu	Thr	Asp	Ala 90	Lys	Leu	Val	His	Glu 95	Ala	
Met	Ile	Gln	Lys 100	Gly	Ala	Thr	Tyr	Ala 105	Thr	Arg	Pro	Pro	Glu 110	Asn	Pro	
Thr	Arg	Thr 115	Ile	Phe	Ser	Glu	Asn 120		Phe	Thr	Val	Asn 125	Ala	Ala	Thr	
Tyr	Gly 130	Pro	Val	Trp	Lys	Ser 135	Leu	Arg	Arg	Asn	Met 140	Val	Gln	Asn	Met	
Leu 145	Ser	Ser	Thr	Arg	Leu 150		Glu	Phe	Arg	Ser 155		Arg	Asp	Asn	Ala 160	
Met	Asp	Lys	Leu	Ile 165	Asn	Arg	Leu	Lys	Asp 170		Ala	Glu	Lys	Asn 175		

PCT/US98/20807 WO 99/19493

									-52	-					
Gly	Val	Val	Trp 180	Val	Leu	Lys	Asp	Ala 185	Arg	Phe	Ala	Val	Phe 190	Cys	Ile
Leu	Val	Ala 1 95	Met	Cys	Phe	Gly	Leu 200	Glu	Met	Asp	Glu	Glu 205	Thr	Val	Glu
Arg	Ile 210	Asp	Gln	Val	Met	Lys 215	Ser	Val	Leu	Ile	Thr 220	Leu	Asp	Prc	Arg
Iie 225	Asp	Asp	Tyr	Leu	Pro 230	Ile	Leu	Ser	Pro	Phe 235	Phe	Ser	Lys	Gln	Arg 240
Lys	Lys	Ala	Leu	Glu 245	Val	Arg	Arg	Glu	Gln 250	Val	Glu	Phe	Leu	Val 255	Pro
Ile	Ile	Glu	Gln 260	Arg	Arg	Arg	Ala	Ile 265	Gln	Asn	Pro	Gly	Ser 270	Asp	His
Thr	Ala	Thr 275	Thr	Phe	Ser	Tyr	Leu 280	Asp	Thr	Leu	Phe	Asp 285	Leu	Lys	Val
Glu	Gly 290	Lys	Lys	Ser	Ala	Pro 295	Ser	Asp	Ala	Glu	Leu 300	Val	Ser	Leu	Cys
Ser 305	Glu	Phe	Leu	Asn	Gly 310	Gly	Thr	Asp	Thr	Thr 315	Ala	Thr	Ala	Val	Glu 320
Trp	Gly	Ile	Ala	Gln 325	Leu	Ile	Ala	Asn	Pro 330	Asn	Val	Gln	Thr	Lys 335	Leu
Tyr	Glu	Ġlu	Ile 340	Lys	Arg	Thr	Val	Gly 345	Glu	Lys	Lys	Val	Asp 350	Glu	Lys
Asp	Val	Glu 355	Lys	Met	Pro	Tyr	Leu 360	His	Ala	Val	Val	Lys 365	Glu	Leu	Leu
Arg	Lys 370	His	Pro	Pro	Thr	His 375	Phe	Val	Leu	Thr	His 380	Ala	Val	Thr	Glu
Pro 385		Thr	Leu	Gly	Gly 390	Tyr	Asp	Ile	Pro	Ile 395	Asp	Ala	Asn	Val	Glu 400
Val	Tyr	Thr	Pro	Ala 405	Ile	Ala	Glu	Asp	Pro 410	Lys	Asn	Trp	Leu	Asn 415	Pro
Glu	Lys	Phe	Asp 420	Pro	Glu	Arg	Phe	Ile 425		Gly	Gly	Glu	Glu 430	Ala	Asp
Ile	Thr	Gly 435	Val	Thr	Gly	Val	Lys 440		Met	Pro	Phe	Gly 445	Val	Gly	Arg
Arg	Ile 450	_	Pro	Gly	Leu	Ala 455	Met	Ala	Thr	Val	His 460	Ile	His	Leu	Met
Met 465		Arg	Met	Val	Gln 470		Phe	Glu	Trp	Gly 475		Tyr	Pro	Pro	Gl: 480
Lys	Lys	Met	Asp	Phe		Gly	Lys	Trp	Glu 490		Thr	Val	Val	Met	

-53-

Glu Ser Leu Arg Ala Thr Ile Lys Pro Arg Gly Gly Glu Lys Val Lys 505 500

Leu

- (2) INFORMATION FOR SEQ ID NO:7:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1611 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: CDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 20..1588
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AAGCACTATC CCTCCCACC ATG ACA AGC CAC ATT GAC GAC AAC CTC TGG ATA Met Thr Ser His Ile Asp Asp Asn Leu Trp Ile 1 5 10	. 52
ATA GCC CTG ACC TCG AAA TGC ACC CAA GAA AAC CTT GCA TGG GTC CTT Ile Ala Leu Thr Ser Lys Cys Thr Gln Glu Asn Leu Ala Trp Val Leu 15 20 25	100
TTG ATC ATG GGC TCA CTC TGG TTA ACC ATG ACT TTC TAT TAC TGG TCA Leu Ile Met Gly Ser Leu Trp Leu Thr Met Thr Phe Tyr Tyr Trp Ser 30 35 40	148
CAC CCC GGT GGT CCT GCC TGG GGC AAG TAC TAC ACC TAC TCT CCC CCC His Pro Gly Gly Pro Ala Trp Gly Lys Tyr Tyr Thr Tyr Ser Pro Pro 45 50 55	196
CTT TCA ATC ATT CCC GGT CCC AAA GGC TTC CCT CTT ATT GGA AGC ATG Leu Ser Ile Ile Pro Gly Pro Lys Gly Phe Pro Leu Ile Gly Ser Met 60 65 70 75	244
GGC CTC ATG ACT TCC CTG GCC CAT CAC CGT ATC GCA GCC GCG GCC GCC Gly Leu Met Thr Ser Leu Ala His His Arg Ile Ala Ala Ala Ala 80 85 90	292
ACA TGC AGA GCC AAG CGC CTC ATG GCC TTT AGT CTC GGC GAC ACA CGT Thr Cys Arg Ala Lys Arg Leu Met Ala Phe Ser Leu Gly Asp Thr Arg 95 100 105	340
GTC ATC GTC ACG TGC CAC CCC GAC GTG GCC AAG GAG ATT CTC AAC AGC Val Ile Val Thr Cys His Pro Asp Val Ala Lys Glu Ile Leu Asn Ser 110 125	388
TCC GTC TTC GCC GAT CGT CCC GTC AAA GAA TCC GCA TAC AGC CTC ATG Ser Val Phe Ala Asp Arg Pro Val Lys Glu Ser Ala Tyr Ser Leu Met 125	436

-54-

TTT Phe 140	AAC Asn	CGC Arg	GCC Ala	ATC Ile	GGC Gly 145	TTC Phe	GCC Ala	TCT Ser	TAC Tyr	GGA Gly 150	GTT Val	TAC Tyr	TGG Trp	CGA Arg	AGC Ser 155	484
CTC Leu	AGG Arg	AGA Arg	ATC Ile	GCC Ala 160	TCT Ser	AAT Asn	CAC His	CTC Leu	TTC Phe 165	TGC Cys	bro CCC	CGC Ang	CAG Gln	ATA Ile 170	AAA Ly:s	532
GCC Ala	TCT Ser	GAG Glu	CTC Leu 175	CAA Gln	CGC Ary	TCT Ser	CAA Gln	ATC Ile 180	GCC Ala	GCC Ala	CAA Gln	ATG Met	GTT Val 185	CAC His	ATC Ile	580
CTA Leu	AAT Asn	AAC Asn 190	AAG Lys	CGC Arg	CAC His	CGC Arg	AGC Ser 195	TTA Leu	CGT Arg	GTT Val	CGC Arg	CAA Gln 200	GTG Val	CTG Leu	AAA Lys	628
AAG Lys	GCT Ala 205	TCG Ser	CTC Leu	AGT Ser	AAC Asn	ATG Met 210	ATG Met	TGC Cys	TCC Ser	GTG Val	TTT Phe 215	GGA Gly	CAA Gln	GAG Glu	TAT Tyr	676
AAG Lys 220	CTG Leu	CAC His	GAC Asp	CCA Pro	AAC Asn 225	AGC Ser	GGA Gly	ATG Met	GAA Glu	GAC Asp 230	CTT Leu	GGA Gly	ATA Ile	TTA Leu	GTG Val 235	724
GAC Asp	CAA Gln	GGT Gly	TAT Tyr	GAC Asp 240	CTG Leu	TTG Leu	GGC Gly	CTG Leu	TTT Phe 245	AAT Asn	TGG Trp	GCC Ala	GAC Asp	CAC His 250	CTT Leu	772
CCT Pro	TTT Phe	CTT Leu	GCA Ala 255	CAT His	TTC Phe	GAC Asp	GCC Ala	CAA Gln 260	AAT Asn	ATC Ile	CGG Arg	TTC Phe	AGG Arg 265	TGC Cys	TCC Ser	820
AAC Asn	CTC Leu	GTC Val 270	CCC Pro	ATG Met	GTG Val	AAC Asn	CGT Arg 275	TTC Phe	GTC Val	GGC	ACA Thr	ATC Ile 280	ATC Ile	GCT Ala	GAA Glu	868
CAC His	CGA Arg 285	GCT Ala	AGT Ser	AAA Lys	ACC Thr	GAA Glu 290	ACC Thr	AAT Asn	CGT Arg	GAT Asp	TTT Phe 295	GTT Val	GAC Asp	GTC Val	TTG Leu	916
CTC Leu 300	TCT Ser	CTC Leu	CCG Pro	GAA Glu	CCT Pro 305	GAT Asp	CAA Gln	TTA Leu	TCA Ser	GAC Asp 310	TCC Ser	GAC Asp	ATG Met	ATC Ile	GCT Ala 315	964
GTA Val	CTT Leu	TGG Trp	GAA Glu	ATG Met 320	ATA Ile	TTC Phe	AGA Arg	GGA Gly	ACG Thr 325	GAC Asp	ACG Thr	GTA Val	GCG Ala	GTT Val 330	TTG Leu	1012
			ATA Ile 335						Leu							1060
AAA Lys	GTT Val	CAA Gln 350	GAG Glu	GAG Glu	CTA Leu	GAT Asp	GCA Ala 355	GTT Val	GTC Val	GGA Gly	AAA Lys	GCA Ala 360	CGC Arg	GCC Ala	GTC Val	1108
GCA Ala	GAG Glu 365	Asp	GAC Asp	GTG Val	GCA Ala	GTG Val 370	Met	ACG Thr	TAC Tyr	CTA Leu	CCA Pro 375	GCG Ala	GTG Val	GTG Val	AAG Lys	1156

-55-

GAG Glu 380	GTG Val	CTG Leu	CGG Arg	CTG Leu	CAC His 385	CCG Pro	CCG Pro	GGC Gly	CCA Pro	CTT Leu 390	CTA Leu	TCA Ser	TGG Trp	GCC Ala	CGC Arg 395	1204
TTG Leu	TCC Ser	ATC. Ile	AAT Asn	GAT Asp 400	ACG Thr	ACC Thr	ATT Ile	GAT Asp	GGG Gly 405	TAT Tyr	CAC His	GTA Val	CCT Pro	GCG Ala 410	GGG Gly	1252
ACC Thr	ACT Thr	GCT Ala	ATG Met 415	GTC Val	AAC Asn	ACG Thr	TGG 1'rp	GCT Ala 420	ATT Ile	TGC Cys	AGG Arg	GAC Asp	CCA Fro 425	CAC His	GTG Val	1300
TGG Trp	AAG Lys	GAC Asp 430	CCA Pro	CTC Leu	GAA Glu	TTT Phe	ATG Met 435	CCC Pro	GAG Glu	AGG Arg	TTT Phe	GTC Val 440	ACT Thr	GCG Ala	GGT Gly	1348
GGA Gly	GAT Asp 445	GCC Ala	GAA Glu	TTT Phe	TCG Ser	ATA Ile 450	CTC Leu	GGG Gly	TCG Ser	GAT Asp	CCA Pro 455	AGA Arg	CTT Leu	GCT Ala	CCA Pro	1396
TTT Phe 460	GGG Gly	TCG Ser	GGT Gly	AGG Arg	AGA Arg 465	GCG Ala	TGC Cys	CCA Pro	GGG Gly	AAG Lys 470	ACT Thr	CTT Leu	GGA Gly	TGG Trp	GCT Ala 475	1444
ACG Thr	GTG Val	AAC Asn	TTT Phe	TGG Trp 480	GTG Val	GCG Ala	TCG Ser	CTC Leu	TTG Leu 485	CAT His	GAG Glu	TTC Phe	GAA Glu	TGG Trp 490	GTA Val	1492
CCG Pro	TCT Ser	GAT Asp	GAG Glu 495	AAG Lys	GGT Gly	GTT Val	GAT Asp	CTG Leu 500	ACG Thr	GAG Glu	GTG Val	CTG Leu	AAG Lys 505	CTC Leu	TCT Ser	1540
AGT Ser	GAA Glu	ATG Met 510	GCT Ala	AAC Asn	CCT Pro	CTC Leu	ACC Thr 515	GTC Val	AAA Lys	GTG Val	CGC Arg	CCC Pro 520	AGG Arg	CGT Arg	GGA Gly	1588
TAAG	GAGA	GAG :	rtgaz	AGCT'	TT T	T										1611

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 523 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Thr Ser His Ile Asp Asp Asn Leu Trp Ile Ile Ala Leu Thr Ser 1 5 10 15

Lys Cys Thr Gln Glu Asn Leu Ala Trp Val Leu Leu Ile Met Gly Ser 20 25 30

Leu Trp Leu Thr Met Thr Phe Tyr Tyr Trp Ser His Pro Gly Gly Pro 35 40 45

-56-

Ala	Trp 50	Gly	Lys	Tyr	Tyr	Thr 55	Tyr	Ser	Pro	Pro	Leu 60	Ser	Ile	Ile	Pro
Gly 65	Pro	Lys	Gly	Phe	Pro 70	Leu	Ile	Gly	Ser	Met 75	Gly	Leu	Met	Thr	Ser 80
Leu	Ala	His	His	Arg 85	Ile	Ala	Ala	Ala	Ala 90	Ala	Thr	Cys	Arg	Ala 95	Lys
Arg	Leu	Met	Ala 100	Phe	Ser	Leu	Gly	Asp 105	Thr	Arg	Vai	Ile	Val 110	Thr	Cys
His	Pro	Asp 115	Val	Ala	Lys	Glu	Ile 120	Leu	Asn	Ser	Ser	Val 125	Phe	Ala	Asp
Arg	Pro 130	Val	Lys	Glu	Ser	Ala 135	Tyr	Ser	Leu	Met	Phe 140	Asn	Arg	Ala	Ile
Gly 145	Phe	Ala	Ser	Tyr	Gly 150	Val	Tyr	Trp	Arg	Ser 155	Leu	Arg	Arg	Ile	Ala 160
Ser	Asn	His	Leu	Phe 165	Cys	Pro	Arg	Gln	Ile 170	Lys	Ala	Ser	Glu	Leu 175	Gln
Arg	Ser	Gln	Ile 180	Ala	Ala	Gln	Met	Val 185	His	Ile	Leu	Asn	Asn 190	Lys	Arg
His	Arg	Ser 195	Leu	Arg	Val	Arg	Gln 200	Val	Leu	Lys	Lys	Ala 205	Ser	Leu	Ser
Asn	Met 210	Met	Cys	Ser	Val	Phe 215	Gly	Gln	Glu	Tyr	Lys 220	Leu	His	Asp	Pro
Asn 225	Ser	Gly	Met	Glu	Asp 230	Leu	Gly	Ile	Leu	Val 235	Asp	Gln	Gly	Tyr	Asp 240
Leu	Leu	Gly	Leu	Phe 245	Asn	Trp	Ala	Asp	His 250		Pro	Phe	Leu	Ala 255	His
Phe	Asp	Ala	Gln 260	Asn	Ile	Arg	Phe	Arg 265	Cys	Ser	Asn	Leu	Val 270	Pro	Met
Val	Asn	Arg 275	Phe	Val	Gly	Thr	11e 280		Ala	Glu	His	Arg 285	Ala	Ser	Lys
Thr	Glu 290		Asn	Arg	Asp	Phe 295		Asp	Val	Leu	Leu 300		Leu	Pro	Glu
Pro 305		Gln	Leu	Ser	Asp 310		Asp	Met	Ile	Ala 315		Leu	Trp	Glu	Met 320
Ile	Phe	Arg	Gly	Thr 325		Thr	· Val	Ala	Val 330		Ile	Glu	Trp	11e 335	Leu
Ala	Arg	Met	Ala 340		His	Pro	His	Val 345		Ser	Lys	Val	Gln 350		Glu
Leu	Asp	Ala 355	Val	Val	Gly	Lys	Ala 360		Ala	Val	Ala	Glu 365		Asp	Va]

-57-

Ala	Val 370	Met	Thr	Tyr	Leu	Pro 375	Ala	Val	Val	Lys	Glu 380	Val	Leu	Arg	Leu		
His 385	Pro	Pro	Gly	Pro	Leu 390	Leu	Ser	Trp	Ala	Arg 395	Leu	Ser	Ile	Asn	Asp 400		
Thr	Thr	Ile	Asp	Gly 405	Tyr	His	Val	Pro	Ala 410	Gly	Thr	Thr	Ala	Met 415	Val		
Asn	Thr	Trp	Ala 420	Ile	Cys	Arg	Asp	Pro 425	His	Val	Trp	Lys	Asp 430	Pro	Leu		
Glu	Phe	Met 435	Pro	Glu	Arg	Phe	Val 440	Thr	Ala	Gly	Gly	Asp 445	Ala	Glu	Phe		
Ser	Ile 450	Leu	Gly	Ser	Asp	Pro 455	Arg	Leu	Ala	Pro	Phe 460	Gly	Ser	Gly	Arg		
Arg 465	Ala	Cys	Pro	Gly	Lys 470	Thr	Leu	Gly	Trp	Ala 475	Thr	Val	Asn	Phe	Trp 480		
Val	Ala	Ser	Leu	Leu 485	His	Glu	Phe	Glu	Trp 490	Val	Pro	Ser	Asp	Glu 495	Lys		
Gly	Val	Asp	Leu 500	Thr	Glu	Val	Leu	Lys 505	Leu	Ser	Ser	Glu	Met 510	Ala	Asn		
Pro	Leu	Thr 515	Val	Lys	Val	Arg	Pro 520	Arg	Arg	Gly							
(2)	INFO	ORMAT	rion	FOR	SEQ	ID 1	10:9:	;								-,	
	(i)	(F (C	A) LE B) TY C) ST	ENGTI (PE : (RANI	i: 17 nucl DEDNI	788 b Leic	STIC ase acio sing ar	pai:	:s							, i	
	(ii)	MOI	LECUI	E TY	PE:	CDNA	4										
	(ix)	(Z		ME/F		CDS 61	.601										
	(xi)	SEC	QUENC	CE DE	ESCRI	PTIC	ON: 5	SEQ I	D NO	0:9:							
GGG:										ln Hi			or Le				47
													GGC Gly				95
													CAA Gln				143

863

WO 99/19493

CTG GTT GAA GGG TGG TTA GAG GAA CAC AAA AGG AAA AGA GCT TTC AAT

Leu Val Glu Gly Trp Leu Glu Glu His Lys Arg Lys Arg Ala Phe Asn

									-59	-						•
				275					280					285		
ATG Met	GAT Asp	GCA Ala	AAA Lys 290	GAA Glu	GAA Glu	CAG Gln	GAT Asp	AAT Asn 295	TTC Phe	ATG Met	GAT Asp	GTC Val	ATG Met 300	CTG Leu	AAT Asn	911
GTT Val	CTG Leu	AAA Lys 305	CAT Asp	GCA Ala	GAG Glu	ATT Ile	TCT Ser 310	GGT Gly	TAT Tyr	GAT Asp	TCA Ser	GAT Asp 315	ACC Thr	ATC Ile	ATC Ile	959
AAG Lys	GCT Ala 320	ACT Thr	TGT Cys	CTG Leu	AAT Asn	CTG Leu 325	ATT Ile	TTA Leu	GCA Ala	GGA Gly	AGC Ser 330	GAC Asp	ACC Thr	ACC Thr	ATG Mec	1007
ATT Ile 335	TCA Ser	CTA Leu	ACA Thr	TGG Trp	GTG Val 340	CTA Leu	TCT Ser	CTG Leu	CTA Leu	CTT Leu 345	AAC Asn	CAT His	CAA Gln	ATG Met	GAA Glu 350	1055
CTA Leu	AAA Lys	AAA Lys	GTC Val	CAA Gln 355	GAT Asp	GAA Glu	TTG Leu	GAC Asp	ACT Thr 360	TAT Tyr	ATT Ile	GGG Gly	AAG Lys	GAC Asp 365	AGG Arg	1103
AAG Lys	GTG Val	GAA Glu	GAA Glu 370	TCT Ser	GAC Asp	ATA Ile	ACC Thr	AAG Lys 375	TTG Leu	GTG Val	TAC Tyr	CTC Leu	CAA Gln 380	GCC Ala	ATT Ile	1151
GTG Val	AAG Lys	GAA Glu 385	ACA Thr	ATG Met	CGG Arg	CTG Leu	TAT Tyr 390	CCA Pro	CCA Pro	AGT Ser	CCT Pro	CTT Leu 395	ATC Ile	ACC Thr	CTT Leu	1199
CGT Arg	GCA Ala 400	GCC Ala	ATG Met	GAA Glu	GAC Asp	TGC Cys 405	ACC Thr	TTC Phe	TCA Ser	GGT Gly	GGC Gly 410	TAT Tyr	CAC His	ATT Ile	CCT Pro	1247
GCT Ala 415	GGG Gly	ACA Thr	CGT Arg	TTA Leu	ATG Met 420	GTG Val	AAT Asn	GCT Ala	TGG Trp	AAG Lys 425	ATC Ile	CAC His	CGG Arg	GAT Asp	GGT Gly 430	.1295
CGT Arg	GTT Val	TGG Trp	AGT Ser	GAT Asp 435	CCT Pro	CAT His	GAT Asp	TTC Phe	AAG Lys 440	CCT Pro	GGA Gly	AGG Arg	TTC Phe	TTG Leu 445	ACA Thr	1343
AGC Ser	CAC His	AAA Lys	GAT Asp 450	GTT Val	GAT Asp	GTG Val	AAG Lys	GGT Gly 455	CAG Gln	AAC Asn	TAT Tyr	GAG Glu	CTC Leu 460	GTC Val	CCT Pro	1391
TTT Phe	GGT Gly	TCT Ser 465	GGA Gly	AGG Arg	AGA Arg	GCA Ala	TGC Cys 470	CCT Pro	GGA Gly	GCC Ala	TCG Ser	CTG Leu 475	GCT Ala	CTG Leu	CGT Arg	1439
GTG Val	GTG Val 480	His	TTG Leu	ACC Thr	ATG Met	GCT Ala 485	AGA Arg	CTG Leu	TTA Leu	CAT His	TCT Ser 490	Phe	AAT Asn	GTT Val	GCT Ala	1487
TCT Ser 495	Pro	TCA Ser	AAT Asn	CAA Gln	GTT Val 500	GTG Val	GAC Asp	ATG Met	ACA Thr	GAG Glu 505	AGC Ser	ATT Ile	GGA Gly	CTC Leu	ACA Thr 510	1535
AAT	TTA	AAA	GCA	ACC	CCG	CTT	GAA	ATT	CTC	CTA	ACT	CCA	CGT	CTA	GAC	1583

-60-

Asn Leu Lys Ala Thr Pro Leu Glu Ile Leu Leu Thr Pro Arg Leu Asp 515 520 525

ACC AAA CTT TAT GAG AAC TAGATTAAAT TAAGCTAGTT TTCTCCCAAA 1631 Thr Lys Leu Tyr Glu Asn

TAAGGGGAGG GGTCCTCTAG GTCCTGAAAT CGGGTAATAA CAATAACATG GTTAATCCAG 1691
CTTCCATGTA GGATAATGAT TATTCACTCA TGGGTCACCT TTTAATGGAG CCTCAGTGTA 1751
TTATAATAAC TCCAAACTTG TGGGTCACAA TCCCCCC 1788

- (2) INFORMATION FOR SEQ ID NO:10:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 532 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Gly Met Ala Met Asp Ala Phe Gln His Gln Thr Leu Ile Ser Ile 1 5 10 15

Ile Leu Ala Met Leu Val Gly Val Leu Ile Tyr Gly Leu Lys Arg Thr 20 25 30

His Ser Gly His Gly Lys Ile Cys Ser Ala Pro Gln Ala Gly Gly Ala 35 40 45

Trp Pro Ile Ile Gly His Leu His Leu Phe Gly Gly His Gln His Thr 50 55 60

His Lys Thr Leu Gly Ile Met Ala Glu Lys His Gly Pro Ile Phe Thr 65 70 75 80

Ile Lys Leu Gly Ser Tyr Lys Val Leu Val Leu Ser Ser Trp Glu Met 85 90 95

Ala Lys Glu Cys Phe Thr Val His Asp Lys Ala Phe Ser Thr Arg Pro 100 105 110

Cys Val Ala Ala Ser Lys Leu Met Gly Tyr Asn Tyr Ala Met Phe Gly
115 120 125

Phe Thr Pro Tyr Gly Pro Tyr Trp Arg Glu Ile Arg Lys Leu Thr Thr

Ile Gln Leu Leu Ser Asn His Arg Leu Glu Leu Leu Lys Asn Thr Arg 145 150 155 160

Thr Ser Glu Ser Glu Val Ala Ile Arg Glu Leu Tyr Lys Leu Trp Ser 165 170 175

Arg Glu Gly Cys Pro Lys Gly Gly Val Leu Val Asp Met Lys Gln Trp 180 185 190

-61-

Phe	Gly	Asp 195	Leu	Thr	His	Asn	11e 200	Val	Leu	Arg	Met	Val 205	Arg	Gly	Lys
Pro	Tyr 210	Tyr	Asp	Gly	Ala	Ser 215	Asp	Asp	Tyr	Ala	Glu 220	Gly	Glu	Ala	Arg
Arg 225	Tyr	Lys	Lys	Val	Met 230	Gly	Glu	Cys	Val	Ser 235	Leu	Phe	Gly	Val	Phe 240
Val	Leu	Ser	qaA	Ala 245	Ile	Pro	Phe	Leu	Gly 250	Trp	Leu	Asp	Ile	Asn 255	Gly
Tyr	Glu	Lys	Ala 260	Met	Lys	Arg	Thr	Ala 265	Ser	Glu	Leu	Asp	Pro 270	Leu	Val
Glu	Gly	Trp 275	Leu	Glu	Glu	His	Lys 280	Arg	Lys	Arg	Ala	Phe 285	Asn	Met	Asp
Ala	Lys 290	Glu	Glu	Gln	Asp	Asn 295	Phe	Met	Asp	Val	Met 300	Leu	Asn	Val	Leu
Lys 305	Asp	Ala	Glu	Ile	Ser 310	Gly	Tyr	Asp	Ser	Asp 315	Thr	Ile	Ile	Lys	Ala 320
Thr	Cys	Leu	Asn	Leu 325	Ile	Leu	Ala	Gly	Ser 330	Asp	Thr	Thr	Met	Ile 335	Ser
Leu	Thr	Trp	Val 340	Leu	Ser	Leu	Leu	Leu 345	Asn	His	Gln	Met	Glu 350	Leu	Lys
Lys	Val	Gln 355	Asp	Glu	Leu	Asp	Thr 360	Tyr	Ile	Gly	Lys	Asp 365	Arg	Lys	Val
Glu	Glu 370	Ser	Asp	Ile	Thr	Lys 375	Leu	Val	Tyr	Leu	Gln 380	Ala	Ile	Val	Lys
Glu 385	Thr	Met	Arg	Leu	Tyr 390	Pro	Pro	Ser	Pro	Leu 395	Ile	Thr	Leu	Arg	Ala 400
Ala	Met	Glu	Asp	Cys 405	Thr	Phe	Ser	Gly	Gly 410	Tyr	His	Ile	Pro	Ala 415	Gly
Thr	Arg	Leu	Met 420	Val	Asn	Ala	Trp	Lys 425	Ile	His	Arg	Asp	Gly 430	Arg	Val
Trp	Ser	Asp 435	Pro	His	Asp	Phe	Lys 440	Pro	Gly	Arg	Phe	Leu 445	Thr	Ser	His
Lys	Asp 450	Val	Asp	Val	Lys	Gly 455	Gln	Asn	Tyr	Glu	Leu 460	Val	Pro	Phe	Gly
Ser 465	Gly	Arg	Arg	Ala	Cys 470	Pro	Gly	Ala	Ser	Leu 475	Ala	Leu	Arg	Val	Val 480
His	Leu	Thr	Met	Ala 485	Arg	Leu	Leu	His	Ser 490	Phe	Asn	Val	Ala	Ser 495	Pro
Ser	Asn	Gln	Val 500	Val	Asp	Met	Thr	Glu 505	Ser	Ile	Gly	Leu	Thr 510	Asn	Leu

PCT/US98/20807 WO 99/19493

-62-

Lys Ala Thr Pro Leu Glu Ile Leu Leu Thr Pro Arg Leu Asp Thr Lys 520 525 515

Leu Tyr Glu Asn 530

- (2) INFORMATION FOR SEQ ID NO:11:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1657 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:

115

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1548
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

(322)	·	2	 	 	_					
 								CTC Leu 15		48
 								CCA Pro		96
 								AAC Asn		144
								CCT Pro		192
 								TCG Ser		240
 								GCT Ala 95	_	288
								TTG Leu		336

ATG GGC TTC GCA CCG TAC GGC CCG TAC TGG AGA GAA ATG AAG AAA CTC Met Gly Phe Ala Pro Tyr Gly Pro Tyr Trp Arg Glu Met Lys Lys Leu

TGC ATC GTT CAC CTC TTC AGC GCG CAA CGC GTT CGG TCC TTT CGA CCA

Cys Ile Val His Leu Phe Ser Ala Gln Arg Val Arg Ser Phe Arg Pro

432

-63-

	130					135			-03	-	140						
ATT Ile 145	CGA Arg	GAG Glu	AAC Asn	GAG Glu	GTT Val 150	GCA Ala	AAA Lys	ATG Met	GTT Val	CGG Arg 155	AAA Lys	CTG Leu	TCG Ser	GAA Glu	CAC His 160	48	30
GAA Glu	GCT Ala	TCG Ser	GGT Gly	ACT Thr 165	GTC Val	GTG Val	AAC Asn	TTG Leu	ACC Thr 170	GAA Glu	ACT Thr	TTG Leu	ATG Met	TCT Ser 175	TTC Phe	52	8.8
ACG Thr	AAC Asn	TCT Ser	TTG Leu 180	ATA Ile	TGC Cys	ACA Arg	ATC Ile	GCC Ala 185	TTG Leu	GGG Cly	AAA Lys	AGT Ser	TAC Tyr 190	GGT Gly	TGT Cys	57	76
GAG Glu	TAC Tyr	GAG Glu 195	GAA Glu	GTA Val	GTT Val	GTT Val	GAT Asp 200	GAG Glu	GTA Val	CTG Leu	GGA Gly	AAC Asn 205	CGG Arg	AGG Arg	AGC Ser	62	24
AGG Arg	TTG Leu 210	CAG Gln	GTT Val	CTG Leu	CTC Leu	AAC Asn 215	GAG Glu	GCT Ala	CAA Gln	GCG Ala	TTG Leu 220	CTT Leu	TCG Ser	GAG Glu	TTT Phe	67	72
TTC Phe 225	TTT Phe	TCG Ser	GAT Asp	TAT Tyr	TTT Phe 230	CCG Pro	CCT Pro	ATA Ile	GGA Gly	AAG Lys 235	TGG Trp	GTT Val	GAT Asp	AGA Arg	GTG Val 240	72 :	20
ACG Thr	GGA Gly	ATT Ile	CTA Leu	TCG Ser 245	CGG Arg	CTT Leu	GAT Asp	AAA Lys	ACG Thr 250	TTC Phe	AAG Lys	GAG Glu	TTG Leu	GAC Asp 255	GCG Ala	76	58
TGC Cys	TAC Tyr	GAA Glu	CGA Arg 260	TCA Ser	TCC Ser	TAT Tyr	GAT Asp	CAC His 265	ATG Met	GAT Asp	TCG Ser	GCA Ala	AAG Lys 270	AGT Ser	GGT Gly	81	16
AAA Lys	AAA Lys	GAT Asp 275	AAT Asn	GAC Asp	AAC Asn	AAA Lys	GAA Glu 280	GTC Val	AAA Lys	GAT Asp	ATT Ile	ATT Ile 285	GAT Asp	ATT Ile	CTT Leu	.8 6	54
CTC Leu	CAG Gln 290	CTA Leu	CTT Leu	GAT Asp	GAT Asp	CGT Arg 295	TCC Ser	TTC Phe	ACC Thr	TTT Phe	GAT Asp 300	CTC Leu	ACT Thr	CTC Leu	GAC Asp	93	12
CAC His 305	ATA Ile	AAA Lys	GCC Ala	GTG Val	CTC Leu 310	ATG Met	AAC Asn	ATC Ile	TTT Phe	ATA Ile 315	GCA Ala	GGA Gly	ACA Thr	GAC Asp	CCG Pro 320	96	60
AGT Ser	TCC Ser	GCG Ala	ACA Thr	ATA Ile 325	GTT Val	TGG Trp	GCA Ala	ATG Met	AAT Asn 330	GCA Ala	CTG Leu	TTG Leu	AAG Lys	AAT Asn 335	CCC Pro	100	80
AAT Asn	GTG Val	ATG Met	AGC Ser 340	Lys	GTT Val	CAA Gln	GGA Gly	GAA Glu 345	Val	AGA Arg	AAT Asn	CTA Leu	TTC Phe 350	GGT Gly	GAC Asp	109	56
AAA Lys	GAT Asp	TTC Phe 355	ATA Ile	AAC Asn	GAA Glu	GAT Asp	GAT Asp 360	Val	GAA Glu	AGC Ser	CTT Leu	CCT Pro 365	TAT Tyr	CTC Leu	AAA Lys	110	04
GCA	GTG	GTG	AAG	GAG	ACA	TTA	AGA	TTA	TTC	CCA	CCT	TCA	CCA	CTA	CTT	11	52

PCT/US98/20807 WO 99/19493

									-64	-						
Ala	Val 370	Val	Lys	Glu	Thr	Leu 375	Arg	Leu	Phe	Pro	Pro 380	Ser	Pro	Leu	Leu	
TTG Leu 385	CCA Pro	AGG Arg	GTA Val	ACA Thr	ATG Met 390	GAA Glu	ACA Thr	TGC Cys	AAC Asn	ATA Ile 395	GAA Glu	GGG Gly	TAC Tyr	GAA Glu	ATT Ile 400	1200
														AGG Arg 415		1248
CCT Pro	GAG Glu	AAT Asn	TGG Trp 420	GAA Glu	GAG Glu	CCT Pro	GAG Glu	ALA Lys 425	TTT Phe	TTC Phe	CCC Pro	GAA Glu	AGG Arg 430	TTC Phe	CTT Leu	1296
														ATC Ile		1344
														ATT Ile		1392
														TGG Trp		1440
														ATG Met 495		1488
														GCA Ala		1536
		ACA Thr 515		TAG	CACA	CGT 1	rggtz	ACAT'	IC A	CTAT	AACA(C AC	AAGA	AAGT		1588
TGA'	TAAT	GAC 7	TTGT	GTAT	GC A	ACTA	rgct	C TA	rgca	CTAT	GCA	CTAT	GTT 3	TATTO	GACCAT	1648
TAA	TTAC'	rg														1657

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 516 amino acids
 - (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Leu Val Leu Leu Ser Leu Leu Ser Ile Val Ile Ser Ile Val Leu Phe 10 1 5

Ile Thr His Thr His Lys Arg Asn Asn Thr Pro Arg Gly Pro Pro Gly 25

-65-

Pro	Pro	Pro 35	Leu	Pro	Leu	Ile	Gly 40	Asn	Leu	His	Gln	Leu 45	His	Asn	Ser
Ser	Pro 50	His	Leu	Cys	Leu	Trp 55	Gln	Leu	Ala	Lys	Leu 60	His	Gly	Pro	Leu
Met 65	Ser	Phe	Arg	Leu	Gly 70	Ala	Val	Gln	Thr	Val 75	Val	Val	Ser	Ser	Ala 80
Arg	Ile	Ala	Glu	Gln 85	Ile	Leu	Lys	Thr	His 90	Asp	Leu	Asn	Phe	Ala 95	Ser
Arg	Pro	Leu	Pha 100	Və.1	Gly	Pro	Arg	Lys 105	Leu	Ser	туг	Asp	Gly 110	Leu	Asp
Met	Gly	Phe 115	Ala	Pro	туг	Gly	Pro 120	Tyr	Trp	Arg	Glu	Met 125	Lys	Lys	Leu
Cys	Ile 130	Val	His	Leu	Phe	Ser 135	Ala	Gln	Arg	Val	Arg 140	Ser	Phe	Arg	Pro
Ile 145	Arg	Glu	Asn	Glu	Val 150	Ala	Lys	Met	Val	Arg 155	Lys	Leu	Ser	Glu	His 160
Glu	Ala	Ser	Gly	Thr 165	Val	Val	Asn	Leu	Thr 170	Glu	Thr	Leu	Met	Ser 175	Phe
Thr	Asn	Ser	Leu 180	Ile	Cys	Arg	Ile	Ala 185	Leu	Gly	Lys	Ser	Tyr 190	Gly	Cys
Glu	Tyr	Glu 195	Glu	Val	Val	Val	Asp 200	Glu	Val	Leu	Gly	Asn 205	Arg	Arg	Ser
Arg	Leu 210	Gln	Val	Leu	Leu	Asn 215	Glu	Ala	Gln	Ala	Leu 220	Leu	Ser	Glu	Phe
Phe 225	Phe	Ser	Asp	Tyr	Phe 230	Pro	Pro	Ile	Gly	Lys 2 35	Trp	Val	Asp	Arg	Val 240
Thr	Gly	Ile	Leu	Ser 245	Arg	Leu	Asp	Lys	Thr 250	Phe	Lys	Glu	Leu	Asp 255	Ala
Cys	Tyr	Glu	Arg 260	Ser	Ser	Tyr	Asp	His 265	Met	Asp	Ser	Ala	Lys 270	Ser	Gl
Lys	Lys	Asp 275	Asn	Asp	Asn	Lys	Glu 280	Val	Lys	Asp	Ile	Ile 285	qzA	Ile	Let
Leu	Gln 290	Leu	Leu	Asp	Asp	Arg 295	Ser	Phe	Thr	Phe	Asp 300	Leu	Thr	Leu	Asp
His 305	Ile	Lys	Ala	Val	Leu 310	Met	Asn	Ile	Phe	Ile 315		Gly	Thr	Asp	Pro 320
Ser	Ser	Ala	Thr	11e 325	Val	Trp	Ala	Met	Asn 330	Ala	Leu	Leu	Lys	Asn 335	Pro
Asn	Val	Met	Ser 340	Lys	Val	Gln	Gly	Glu 345		Arg	Asn	Leu	Phe 350	Gly	Ası

-66-

Lys	Asp	Phe 355	Ile	Asn	Glu	Asp	Asp 360	Val	Glu	Ser	Leu	Pro 365	Tyr	Leu	Lys	
Ala	Val 370	Val	Lys	Glu	Thr	Leu 375	Arg	Leu	Phe	Pro	Pro 380	Ser	Pro	Leu	Leu	
Leu 385	Pro	Arg	Val	Thr	Met 390	Glu	Thr	Cys	Asıı	Ile 395	Glu	Gly	Tyr	Glu	11e 400	
Gln	Ala	Lys	Thr	Ile 405	Val	His	Val	Asn	Ala 410	Trp	Ala	Ile	Ala	Arg 415	Asp	
Pro	Glu	Asn	Trp 420	Glu	Glu	Pro	Glu	Lys 425	Phe	Phe	Pro	Glu	Arg 430	Phe	Leu	
Glu	Ser	Ser 435	Met	Glu	Leu	Lys	Gly 440	Asn	Asp	Glu	Phe	Lys 445	Val	Ile	Pro	
Phe	Gly 450	Ser	Gly	Arg	Arg	Met 455	Cys	Pro	Ala	Lys	His 460	Met	Gly	Ile	Met	
Asn 465	Val	Glu	Leu	Ser	Leu 470	Ala	Asn	Leu	Ile	His 475	Thr	Phe	Asp	Trp	Glu 480	
Val	Ala	Lys	Gly	Phe 485	Asp	Lys	Glu	Glu	Met 490	Leu	Asp	Thr	Gln	Met 495	Lys	
Pro	Gly	Ile	Thr 500	Met	His	Lys	Lys	Ser 505	Asp	Leu	Tyr	Leu	Val 510	Ala	Lys	
Lys Pro Thr Thr 515																
(2)	INF	ORMA	TION	FOR	SEQ	ID 1	10:1	3:								
(2) INFORMATION FOR SEQ ID NO:13: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1824 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA																
	(ix	(.		AME/	KEY: ION:		.161	6								
	·(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:13	:					
GGA	TAAA	TAG	CCTC	ACAA	AA G	CAAA	GATC	A AA	CAAA	CCAA	GGA	CGAG	AAC		ATG Met 1	56
TTG Leu	CTT Leu	GAA Glu	CTT Leu 5	Ala	CTT Leu	GGT Gly	TTA Leu	TTG Leu 10	Val	TTG Leu	GCT Ala	CTG Leu	TTT Phe 15	Leu	CAC His	104

-67-

									-67								•
TTG Leu	CGT Arg	CCC Pro 20	ACA Thr	CCC Pro	ACT Thr	GCA Ala	AAA Lys 25	TCA Ser	AAA Lys	GCA Ala	CTT Leu	CGC Arg 30	CAT His	CTC Leu	CCA Pro		152
AAC Asn	CCA Pro 35	CCA Pro	AGC Ser	CCA Pro	AAG Lys	CCT Pro 40	CGT Arg	CTT Leu	CCC Pro	TTC Phe	ATA Ile 45	GGA Gly	CAC His	CTT Leu	CAT His		200
CTC Leu 50	TTA Leu	AAA Lys	GAC Asp	AAA Lys	CTT Leu 55	CTC Leu	CAC His	TAC Tyr	GCA Ala	CTC Leu 60	ATC Ile	GAC Asp	CTC Leu	TCC Ser	AAA Lys 65		248
AAA Lys	CAT His	GGT Gly	CCC Pro	TTA Leu 70	TTC Phe	TCT Ser	CTC Leu	TAC Tyr	TTT Phe 75	GGC Gly	TCC Ser	ATG Met	CCA Pro	ACC Thr 80	GTT Val		296
GTT Val	GCC Ala	TCC Ser	ACA Thr 85	CCA Pro	GAA Glu	TTG Leu	TTC Phe	AAG Lys 90	CTC Leu	TTC Phe	CTC Leu	CAA Gln	ACG Thr 95	CAC His	GAG Glu		344
GCA Ala	ACT Thr	TCC Ser 100	TTC Phe	AAC Asn	ACA Thr	AGG Arg	TTC Phe 105	CAA Gln	ACC Thr	TCA Ser	GCC Ala	ATA Ile 110	AGA Arg	CGC Arg	CTC Leu		392
ACC Thr	TAT Tyr 115	GAT Asp	AGC Ser	TCA Ser	GTG Val	GCC Ala 120	ATG Met	GTT Val	CCC Pro	TTC Phe	GGA Gly 125	CCT Pro	TAC Tyr	TGG Trp	AAG Lys		440
TTC Phe 130	GTG Val	AGG Arg	AAG Lys	CTC Leu	ATC Ile 135	ATG Met	AAC Asn	GAC Asp	CTT Leu	CCC Pro 140	AAC Asn	GCC Ala	ACC Thr	ACT Thr	GTA Val 145		488
AAC Asn	AAG Lys	TTG Leu	AGG Arg	CCT Pro 150	TTG Leu	AGG Arg	ACC Thr	CAA Gln	CAG Gln 155	ACC Thr	CGC Arg	AAG Lys	TTC Phe	CTT Leu 160	AGG Arg	.7.	536
GTT Val	ATG Met	GCC Ala	CAA Gln 165	GGC Gly	GCA Ala	GAG Glu	GCA Ala	CAG Gln 170	AAG Lys	CCC Pro	CTT Leu	GAC Asp	TTG Leu 175	ACC Thr	GAG Glu		584
GAG Glu	CTT Leu	CTG Leu 180	AAA Lys	TGG Trp	ACC Thr	AAC Asn	AGC Ser 185	ACC Thr	ATC Ile	TCC Ser	ATG Met	ATG Met 190	ATG Met	CTC Leu	GGC Gly		632
GAG Glu	GCT Ala 195	GAG Glu	GAG Glu	ATC Ile	AGA Arg	GAC Asp 200	ATC Ile	GCT Ala	CGC Arg	GAG Glu	GTT Val 205	CTT Leu	AAG Lys	ATC Ile	TTT Phe		680
GGC Gly 210	GAA Glu	TAC Tyr	AGC Ser	CTC Leu	ACT Thr 215	GAC Asp	TTC Phe	ATC Ile	TGG Trp	CCA Pro 220	TTG Leu	AAG Lys	CAT His	CTC Leu	AAG Lys 225		728
GTT Val	GGA Gly	AAG Lys	TAT Tyr	GAG Glu 230	AAG Lys	AGG Arg	ATC Ile	GAC Asp	GAC Asp 235	ATC Ile	TTG Leu	AAC Asn	AAG Lys	TTC Phe 240	GAC Asp		776
CCT Pro	GTC Val	GTT Val	GAA Glu 245	Arg	GTC Val	ATC Ile	AAG Lys	AAG Lys 250	Arg	CGT Arg	GAG Glu	ATC Ile	GTG Val 255	AGG Arg	AGG Arg		824

-68-

	-00 -												• •			
AGA Arg	AAG Lys	AAC Asn 260	GGA Gly	GAG Glu	GTT Val	GTT Val	GAG Glu 265	GGT Gly	GAG Glu	GTC Val	AGC Ser	GGG Gly 270	GTT Val	TTC Phe	CTT Leu	872
			CTT Leu													920
			CAC His													968
			ACA Thr													1016
			AAG Lys 325													1064
			GAC Asp													1112
			GCA Ala													1160
			AAA Lys													1208
			GAG Glu													1256
			AAA Lys 405													1304
TTC Phe	CTA Leu	GAG Glu 420	ACA Thr	GGG Gly	GCT Ala	GAA Glu	GGG Gly 425	GAA Glu	GCA Ala	GGG Gly	CCT Pro	CTT Leu 430	GAT Asp	CTT Leu	AGG Arg	1352
			TTT Phe													1400
			AAT Asn													1448
			TGC Cys													1496
			GGT Gly 485													1544

-69-

-09-																
ACT GTI Thr Val	CCA Pro 500	AGG Arg	GCA Ala	CAT His	AGT Ser	CTT Leu 505	GTC Val	TGT Cys	GTT Val	CCA Pro	CTT Leu 510	GCA Ala	AGG Arg	AT Il	rc Le	1592
GGC GTT Gly Val 515	Ala						TAAT	AATT	GAT (CATC	ATCA:	FA T.	ATAA	ΓAΊ	ГТТ	1646
ACTTTTT	rgīg :	rgtto	SATA	AT CA	TCAT	TTC	A ATA	\AGGʻ	TCTC	GTT	CATC	rac	TTTT	ΓAΊ	rgaa	1706
GTATATA	LAGC (CCTT	CCAT	SC AC	CATTO	TATO	C ATC	CTCC	CATT	TGT	CTTC	GTT '	TGCT	ACC	CTAA	1766
GGCAATO	TTT :	CTTT	rttt/	AG AA	ATCAC	CATC	A TCC	CTAC'	TATA	AACT	ratc:	AAT	CCTT	ATA	AT	1824

(2) INFORMATION FOR SEQ ID NO.14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 521 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met Leu Leu Glu Leu Ala Leu Gly Leu Leu Val Leu Ala Leu Phe Leu 1 5 10 15

His Leu Arg Pro Thr Pro Thr Ala Lys Ser Lys Ala Leu Arg His Leu 20 25 30

Pro Asn Pro Pro Ser Pro Lys Pro Arg Leu Pro Phe Ile Gly His Leu 35 40 45

His Leu Leu Lys Asp Lys Leu Leu His Tyr Ala Leu Ile Asp Leu Ser 50 55 60

Lys Lys His Gly Pro Leu Phe Ser Leu Tyr Phe Gly Ser Met Pro Thr 65 70 75 80

Val Val Ala Ser Thr Pro Glu Leu Phe Lys Leu Phe Leu Gln Thr His 85 90 95

Glu Ala Thr Ser Phe Asn Thr Arg Phe Gln Thr Ser Ala Ile Arg Arg
100 105 110

Leu Thr Tyr Asp Ser Ser Val Ala Met Val Pro Phe Gly Pro Tyr Trp 115 120 125

Lys Phe Val Arg Lys Leu Ile Met Asn Asp Leu Pro Asn Ala Thr Thr 130 135 140

Val Asn Lys Leu Arg Pro Leu Arg Thr Gln Gln Thr Arg Lys Phe Leu 145 150 155 160

Arg Val Met Ala Gln Gly Ala Glu Ala Gln Lys Pro Leu Asp Leu Thr 165 170 175

Glu Glu Leu Leu Lys Trp Thr Asn Ser Thr Ile Ser Met Met Leu

-70-

									-70	-					
			180					185					190		
Gly	Glu	Ala 195	Glu	Glu	Ile	Arg	Asp 200	Ile	Ala	Arg	Glu	Val 205	Leu	Lys	Ile
Phe	Gly 210	Glu	Tyr	Ser	Leu	Thr 215	Asp	Phe	Ile	Trp	Pro 220	Leu	Lys	His	Leu
Lys 225	Val	Gly	Lys	Tyr	Glu 230	Lys	Arg	Ile	Asp	Asp 235	Ile	Leu	Asn	Lys	Phe 240
Asp	.Pro	Val	Val	Glu 245	Arg	Val	Ile	Lys	Lys 250	Arg	Arg	Glu	rle	Val 255	Arg
Arg	Arg	Lys	Asn 260	Gly	Glu	Val	Val	Glu 265	Gly	Glu	Val	Ser	Gly 270	Val	Phe
Leu	Asp	Thr 275	Leu	Leu	Glu	Phe	Ala 280	Glu	Asp	Glu	Thr	Met 285	Glu	Ile	Lys
Ile	Thr 290	Lys	Asp	His	Ile	Glu 295	Gly	Leu	Val	Val	Asp 300	Phe	Phe	Ser	Ala
Glÿ 305	Thr	Asp	Ser	Thr	Ala 310	Val	Ala	Thr	Glu	Trp 315	Ala	Leu	Ala	Glu	Leu 320
Ile	Asn	Asn	Pro	Lys 325	Val	Leu	Glu	Lys	Ala 330	Arg	Glu	Glu	Val	Tyr 335	Ser
Val	Val	Gly	Lys 340	Asp	Arg	Leu	Val	Asp 345	Glu	Val	Asp	Thr	Gln 350	Asn	Leu
Pro	Tyr	Ile 355	Arg	Ala	Ile	Val	Lys 360	Glu	Thr	Phe	Arg	Met 365	His	Pro	Pro
Leu	Pro 370	Val	Val	Lys	Arg	Lys 375	Cys	Thr	Glu	Glu	Cys 380	Glu	Ile	Asn	Gly
385			Pro		390					395					400
Gly	Arg	Asp	Pro	Lys 405	Tyr	Trp	Asp	Arg	Pro 410		Glu	Phe	Arg	Pro 415	Glu
Arg	Phe	Leu	Glu 420	Thr	Gly	Ala	Glu	Gly 425	Glu	Ala	Gly	Pro	Leu 430	Asp	Leu
Arg	Gly	Gln 435	His	Phe	Gln	Leu	Leu 440		Phe	Gly	Ser	Gly 445	Arg	Arg	Met
Cys	Pro 450	_	Val	Asn	Leu	Ala 455		Ser	Gly	Met	Ala 460		Leu	Leu	Ala
Ser 465		Ile	Gln	Cys	Phe 470		Leu	Gln	Val	Leu 475		Pro	Gln	Gly	Gln 480
Ile	Leu	Lys	Gly	Gly 485		Ala	Lys	Val	Ser 490		Glu	Glu	Arg	Ala 495	Gly
Leu	Thr	Val	Pro	Arg	Ala	His	Ser	Leu	Val	Cys	Val	Pro	Leu	Ala	Arg

WO 99/19493 PCT/US98/20807

-71-

500 505 510

Ile Gly Val Ala Ser Lys Leu Leu Ser 515 520

- (2) INFORMATION FOR SEQ ID NO:15:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1831 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 20..1747
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CAACACTCGC AGTACCGCC ATG AGT GTC GAC ACT TCC TCC ACC CTC TCC ACC Met Ser Val Asp Thr Ser Ser Thr Leu Ser Thr 1 5 10	52
GTC ACC GAT GCC AAT CTT CAC TCC AGA TTT CAT TCT CGT CTT GTT CCA Val Thr Asp Ala Asn Leu His Ser Arg Phe His Ser Arg Leu Val Pro 15 20 25	100
TTC ACT CAT CAT TTC TCA CTT TCT CAA CCC AAA CGG ATT TCT TCA ATC Phe Thr His His Phe Ser Leu Ser Gln Pro Lys Arg Ile Ser Ser Ile 30 35 40	148
AGA TGC CAA TCA ATT AAT ACC GAT AAG AAG AAA TCA AGT AGA AAT CTG Arg Cys Gln Ser Ile Asn Thr Asp Lys Lys Ser Ser Arg Asn Leu 45 50 55	_. 196
CTG GGC AAT GCA AGT AAC CTC CTC ACG GAC TTA TTA AGT GGT GGA AGT Leu Gly Asn Ala Ser Asn Leu Leu Thr Asp Leu Leu Ser Gly Gly Ser 60 65 70 75	244
ATA GGG TCT ATG CCC ATA GCT GAA GGT GCA GTC TCA GAT CTG CTT GGT Ile Gly Ser Met Pro Ile Ala Glu Gly Ala Val Ser Asp Leu Leu Gly 80 85 90	292
CGA CCT CTC TTT TTC TCA CTG TAT GAT TGG TTC TTG GAG CAT GGT GCG Arg Pro Leu Phe Phe Ser Leu Tyr Asp Trp Phe Leu Glu His Gly Ala 95 100 105	340
GTG TAT AAA CTT GCC TTT GGA CCA AAA GCA TTT GTT GTT GTA TCA GAT Val Tyr Lys Leu Ala Phe Gly Pro Lys Ala Phe Val Val Val Ser Asp 110 115 120	388
CCC ATA GTT GCT AGA CAT ATT CTG CGA GAA AAT GCA TTT TCT TAT GAC Pro Ile Val Ala Arg His Ile Leu Arg Glu Asn Ala Phe Ser Tyr Asp 125 130 135	436
AAG GGA GTA CTT GCT GAT ATC CTT GAA CCA ATA ATG GGC AAA GGA CTC	484

_3

									-72	_							
Lys 140	Gly	Val	Leu	Ala	Asp 145	Ile	Leu	Glu	Pro	Ile 150	Met	Gly	Lys	Gly	Leu 155		
ATA Ile	CCA Pro	GCA Ala	GAC Asp	CTT Leu 160	GAT Asp	ACT Thr	TGG Trp	AAG Lys	CAA Gln 165	AGG Arg	AGA Arg	AGA Arg	GTC Val	ATT Ile 170	GCT Ala		532
CCG Pro	GCT Ala	TTC Phe	CAT His 175	AAC Asn	TCA Ser	TAC Tyr	TTG Leu	GAA Glu 180	GCT Ala	ATG Met	GTT Val	AAA Lys	ATA Ile 185	TTC Phe	ACA Thr		580
ACT Thr	TGT Cys	TCA Ser 190	GAA Glu	AGA Arg	ACA Thr	ATA Ile	TTG Leu 195	AAG Lys	TTT Phe	AAT Asn	AAG Lys	CTT Leu 200	CTT Leu	GAA. Glu	GGA Gly	٠	628
GAG Glu	GGT Gly 205	TAT Tyr	GAT Asp	GGA Gly	CCT Pro	GAC Asp 210	TCA Ser	ATT Ile	GAA Glu	TTG Leu	GAT Asp 215	CTT Leu	GAG Glu	GCA Ala	GAG Glu		676
TTT Phe 220	TCT Ser	AGT Ser	TTG Leu	GCT Ala	CTT Leu 225	GAT Asp	ATT Ile	ATT Ile	GGG Gly	CTT Leu 230	GGT Gly	GTG Val	TTC Phe	AAC Asn	TAT Tyr 235		724
GAC Asp	TTT Phe	GGT Gly	TCT Ser	GTC Val 240	ACC Thr	AAA Lys	GAA Glu	TCT	CCA Pro 245	GTT Val	ATT Ile	AAG Lys	GCA Ala	GTC Val 250	TAT Tyr		772
GGC Gly	ACT Thr	CTT Leu	TTT Phe 255	GAA Glu	GCT Ala	GAA Glu	CAC His	AGA Arg 260	TCC Ser	ACT Thr	TTC Phe	TAC Tyr	ATT Ile 265	CCA Pro	TAT Tyr		820
TGG Trp	AAA Lys	ATT Ile 270	CCA Pro	TTG Leu	GCA Ala	AGG Arg	TGG Trp 275	ATA Ile	GTC Val	CCA Pro	AGG Arg	CAA Gln 280	AGA Arg	AAG Lys	TTT Phe		868
CAG Gln	GAT Asp 285	GAC Asp	CTA Leu	AAG Lys	GTC Val	ATC Ile 290	AAT Asn	ACT Thr	TGT Cys	CTT	GAT Asp 295	GGA Gly	CTT	ATC Ile	AGA Arg		916
AAT Asn 300	GCA Ala	AAA Lys	GAG Glu	AGC Ser	AGA Arg 305	CAG Gln	GAA Glu	ACA Thr	GAT Asp	GTT Val 310	GAG Glu	AAA Lys	TTG Leu	CAG Gln	CAG Gln 315		964
AGG Arg	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 320	TTG Leu	AAG Lys	GAT Asp	GCA Ala	AGT Ser 325	Leu	CTG Leu	CGT Arg	TTC Phe	CTG Leu 330	GTT Val		1012
GAT Asp	ATG Met	CGG Arg	GGA Gly 335	Ala	GAT Asp	GTT Val	GAT Asp	GAT Asp 340	Arg	CAG Gln	TTG Leu	AGG Arg	GAT Asp 345	GAT Asp	TTA Leu		1060
ATG Met	ACA Thr	ATG Met 350	Leu	ATT	GCC Ala	GGT Gly	CAT His	Glu	ACA Thr	ACG Thr	GCT Ala	GCA Ala 360	Val	CTT Leu	ACT Thr		1108
TGG Trp	GCA Ala 365	Val	TTC Phe	CTC Leu	CTA Leu	GCT Ala 370	Gln	AAT Asn	CCT	AGC Ser	Lys 375	Met	AAG Lys	AAG Lys	GCT Ala		1156

WO 99/19493 PCT/US98/20807

110 //1/17473	
-73-	1204
CAA GCA GAG GTA GAT TTG GTG CTG GGT ACG GGG AGG CCA ACT TTT GAA Gln Ala Glu Val Asp Leu Val Leu Gly Thr Gly Arg Pro Thr Phe Glu 380 395	1204
TCA CTT AAG GAA TTG CAG TAC ATT AGA TTG ATT GTT GTG GAG GCT CTT Ser Leu Lys Glu Leu Gln Tyr lie Arg Leu Ile Val Val Glu Ala Leu 400 405 410	1252
CGT TTA TAC CCC CAA CCA CCT TTG CTG ATT AGA CGT TCA CTC AAA TCT Arg Leu Tyr Pro Gln Pro Pro Leu Leu Ile Arg Arg Ser Leu Lys Ser 415 420 425	1300
GAT GTT TTA CCA GGT GGG CAC AAA GGT GAA AAA GAT GGT TAT GCA ATT Asp Val Leu Pro Gly Gly His Lys Gly Glu Lys Asp Gly Tyr Ala Ile 430 435	1348
CCT GCT GGG ACT GAT GTC TTC ATT TCT GTA TAT AAT CTC CAT AGA TCT Pro Ala Gly Thr Asp Val Phe Ile Ser Val Tyr Asn Leu His Arg Ser 445 450 455	1396
CCA TAT TTT TGG GAC CGC CCT GAT GAC TTC GAA CCA GAG AGA TTT CTT Pro Tyr Phe Trp Asp Arg Pro Asp Asp Phe Glu Pro Glu Arg Phe Leu 460 475	1444
GTG CAA AAC AAG AAT GAA GAA ATT GAA GGA TGG GCT GGT CTT GAT CCA Val Gln Asn Lys Asn Glu Glu Ile Glu Gly Trp Ala Gly Leu Asp Pro 480 485 490	1492
TCT CGA AGT CCC GGA GCC TTG TAT CCG AAC GAG GTT ATA TCG GAT TTT Ser Arg Ser Pro Gly Ala Leu Tyr Pro Asn Glu Val Ile Ser Asp Phe 495 500 505	1540
GCA TTC TTA CCT TTT GGT GGC GGA CCA CGA AAA TGT GTT GGG GAC CAA Ala Phe Leu Pro Phe Gly Gly Gly Pro Arg Lys Cys Val Gly Asp Gln 510 515 520	1588
TTT GCT CTG ATG GAG TCC ACT GTA GCG TTG ACT ATG CTG CTC CAG AAT Phe Ala Leu Met Glu Ser Thr Val Ala Leu Thr Met Leu Leu Gln Asn 525 530 535	1636
TTT GAC GTG GAA CTA AAA GGG ACC CCT GAA TCG GTG GAA CTA GTT ACT Phe Asp Val Glu Leu Lys Gly Thr Pro Glu Ser Val Glu Leu Val Thr 540 545 550 555	1684
GGG GCA ACT ATT CAT ACC AAA AAT GGA ATG TGG TGC AGA TTG AAG AAG Gly Ala Thr Ile His Thr Lys Asn Gly Met Trp Cys Arg Leu Lys Lys 560 565 570	1732
AGA TCT AAT TTA CGT TGACATATGT ACTGTGGCCA TTTTTCTTAT ACAGAATAAT Arg Ser Asn Leu Arg 575	1787
GTATATTATT ATTCTTTGAG AATAATATGA ATAAATTCCT AGAC	1831

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 576 amino acids

WO 99/19493 PCT/US98/20807

-74-

- (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Ser Val Asp Thr Ser Ser Thr Leu Ser Thr Val Thr Asp Ala Asn
1 5 10 15

Leu His Ser Arg Phe His Ser Arg Leu Val Pro Fhe Thr His His Phe 20 25 30

Ser Leu Ser Gln Pro Lys Arg Ile Ser Ser Ile Arg Cys Gln Ser Ile 35 40 45

Asn Thr Asp Lys Lys Ser Ser Arg Asn Leu Leu Gly Asn Ala Ser

Asn Leu Leu Thr Asp Leu Leu Ser Gly Gly Ser Ile Gly Ser Met Pro 65 70 75 80

Ile Ala Glu Gly Ala Val Ser Asp Leu Leu Gly Arg Pro Leu Phe Phe 85 90 95

Ser Leu Tyr Asp Trp Phe Leu Glu His Gly Ala Val Tyr Lys Leu Ala

Phe Gly Pro Lys Ala Phe Val Val Val Ser Asp Pro Ile Val Ala Arg 115 120 125

His Ile Leu Arg Glu Asn Ala Phe Ser Tyr Asp Lys Gly Val Leu Ala 130 135 140

Asp Ile Leu Glu Pro Ile Met Gly Lys Gly Leu Ile Pro Ala Asp Leu 145 150 155 160

Asp Thr Trp Lys Gln Arg Arg Arg Val Ile Ala Pro Ala Phe His Asn 165 170 175

Ser Tyr Leu Glu Ala Met Val Lys Ile Phe Thr Thr Cys Ser Glu Arg 180 185 190

Thr Ile Leu Lys Phe Asn Lys Leu Leu Glu Gly Glu Gly Tyr Asp Gly
195 200 205

Pro Asp Ser Ile Glu Leu Asp Leu Glu Ala Glu Phe Ser Ser Leu Ala 210 215 220

Leu Asp Ile Ile Gly Leu Gly Val Phe Asn Tyr Asp Phe Gly Ser Val 225 230 235 240

Thr Lys Glu Ser Pro Val Ile Lys Ala Val Tyr Gly Thr Leu Phe Glu 245 250 255

Ala Glu His Arg Ser Thr Phe Tyr Ile Pro Tyr Trp Lys Ile Pro Leu 260 265 270

Ala Arg Trp Ile Val Pro Arg Gln Arg Lys Phe Gln Asp Asp Leu Lys

-75-

- Val Ile Asn Thr Cys Leu Asp Gly Leu Ile Arg Asn Ala Lys Glu Ser 300 Arg Gln Glu Thr Asp Val Glu Lys Leu Gln Gln Arg Asp Tyr Leu Asn 315 310 Leu Lys Asp Ala Ser Leu Leu Arg Phe Leu Val Asp Met Arg Gly Ala 330 Asp Val Asp Asp Arg Gln Leu Arg Asp Asp Leu Met Thr Met Leu Ile Ala Gly His Glu Thr Thr Ala Ala Val Leu Thr Trp Ala Val Phe Leu Leu Ala Gln Asn Pro Ser Lys Met Lys Lys Ala Gln Ala Glu Val Asp 375 Leu Val Leu Gly Thr Gly Arg Pro Thr Phe Glu Ser Leu Lys Glu Leu Gln Tyr Ile Arg Leu Ile Val Val Glu Ala Leu Arg Leu Tyr Pro Gln 410 Pro Pro Leu Leu Ile Arg Arg Ser Leu Lys Ser Asp Val Leu Pro Gly Gly His Lys Gly Glu Lys Asp Gly Tyr Ala Ile Pro Ala Gly Thr Asp Val Phe Ile Ser Val Tyr Asn Leu His Arg Ser Pro Tyr Phe Trp Asp Arg Pro Asp Asp Phe Glu Pro Glu Arg Phe Leu Val Gln Asn Lys Asn 475 470 Glu Glu Ile Glu Gly Trp Ala Gly Leu Asp Pro Ser Arg Ser Pro Gly 490 Ala Leu Tyr Pro Asn Glu Val Ile Ser Asp Phe Ala Phe Leu Pro Phe 505 Gly Gly Gly Pro Arg Lys Cys Val Gly Asp Gln Phe Ala Leu Met Glu 520 Ser Thr Val Ala Leu Thr Met Leu Leu Gln Asn Phe Asp Val Glu Leu 540 535 Lys Gly Thr Pro Glu Ser Val Glu Leu Val Thr Gly Ala Thr Ile His 555 550 545 Thr Lys Asn Gly Met Trp Cys Arg Leu Lys Lys Arg Ser Asn Leu Arg 570
- (2) INFORMATION FOR SEQ ID NO:17:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1704 base pairs

WO 99/19493 PCT/US98/20807

-76-

	TYPE: nucleic	
	STRANDEDNESS:	
(D)	TOPOLOGY: line	ar

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 38..1564

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CAGG	CTCC	CAC A	LAAAC	ATCT	C AT	'CATT	CACC	CAA	CAAP	ATG Met	. Ala	CTG Leu	CTI Leu	CTG Leu 5	ATA Ile	55
ATT Ile	CCC Pro	ATC Ile	TCA Ser 10	CTG Leu	GTC Val	ACC Thr	CTC Leu	TGG Trp 15	CTC Leu	GGT Gly	TAC Tyr	ACC Thr	CTA Leu 20	TAC Tyr	CAG Gln	103
CGA Arg	TTA Leu	AGA Arg 25	TTC Phe	AAG Lys	CTC Leu	CCT Pro	CCG Pro 30	GGT Gly	CCA Pro	CGG Arg	CCC Pro	TGG Trp 35	CCG Pro	GTA Val	GTC Val	151
GGT Gly	AAC Asn 40	CTC Leu	TAC Tyr	GAC Asp	ATA Ile	AAA Lys 45	CCC Pro	GTC Val	CGC Arg	TTC Phe	CGG Arg 50	TGC Cys	TTC Phe	GCG Ala	GAG Glu	199
TGG Trp 55	GCG Ala	CAG Gln	TCT Ser	TAC Tyr	GGC Gly 60	CCC Pro	ATA Ile	ATA Ile	TCG Ser	GTT Val 65	TGG Trp	TTC Phe	GGT Gly	TCG Ser	ACC Thr 70	247
CTA Leu	AAC Asn	GTC Val	ATC Ile	GTT Val 75	TCG Ser	AAC Asn	TCG Ser	GAG Glu	CTG Leu 80	GCG Ala	AAG Lys	GAG Glu	GTG Val	CTG Leu 85	AAG Lys	295
GAG Glu	CAC His	GAT Asp	CAG Gln 90	CTG Leu	CTG Leu	GCG Ala	GAC Asp	CGC Arg 95	CAC His	CGG Arg	AGC Ser	CGG Arg	TCG Ser 100	GCG Ala	GCG Ala	343
AAG Lys	TTC Phe	AGC Ser 105	Arg	GAC Asp	GGG Gly	AAG Lys	GAT Asp 110	CTA Leu	ATT Ile	TGG Trp	GCC Ala	GAT Asp 115	TAT Tyr	GGG Gly	CCG Pro	391
CAC His	TAC Tyr 120	GTG Val	AAG Lys	GTG Val	AGG Arg	AAG Lys 125	GTT Val	TGC Cys	ACG Thr	CTC Leu	GAG Glu 130	CTT Leu	TTC Phe	TCG Ser	CCG Pro	439
AAG Lys 135	Arg	CTC Leu	GAG Glu	GCC Ala	CTG Leu 140	Arg	CCC	ATT Ile	AGG Arg	GAG Glu 145	GAC Asp	GAG Glu	GTC Val	ACC Thr	TCC Ser 150	487
ATG Met	GTT Val	GAC Asp	TCC Ser	GTT Val 155	TAC Tyr	AAT Asn	CAC His	TGC Cys	ACC Thr 160	Ser	ACT Thr	GAA Glu	AAT Asn	TTG Leu 165	GGG Gly	535
AAA Lys	. GGA Gly	ATA Ile	TTG Leu	TTG Leu	AGG Arg	AAG Lys	CAC His	TTG Leu	GGG Gly	GTT Val	GTG Val	GCA Ala	TTC Phe	AAC Asn	AAC Asn	583

				-77-	-						•
	170		175	5				180			
ATA ACC AGG Ile Thr Arg 185	Leu Ala	TTT GGG Phe Gly	AAA AGA Lys Arg 190	TTT phe	GTG . Val .	Asn	TCA Sér 195	GAA Glu	GGT Gly	GTG Val	631
ATG GAT GAG Met Asp Glu 200	CAA GGA Gln Gly	GTA GAA Val Glu 205	TTC AAC	GCC Ala	Ile	GTG Val 210	GAA Glu	AAT Asn	GGG Gly	TTA Leu	679
AAG CTA GGA Lys Leu Gly 215	GCA TCT Ala Ser	CTA GCC Leu Ala 220	ATG GCA	A GAA a Glu	CAC His 225	ATC Ile	CCT Pro	TGG Trp	CTT Leu	CGC Arg 230	727
TGG ATG TTC Trp Met Phe	CCA CTG Pro Leu 235	Glu Glu	GGA GC1 Gly Ala	TTTT a Phe 240	GCC Ala	AAG Lys	CAT His	GGA Gly	GCC Ala 245	Arg CGC	775
CGC GAC CGA Arg Asp Arg	CTC ACC Leu Thr 250	AGA GCC Arg Ala	ATC ATC Ile Met 25	. Ala	GAG Glu	CAC His	ACT Thr	GAA Glu 260	GCA Ala	CGC Arg	823
AAG AAA TCT Lys Lys Ser 265	Gly Gly	GCC AAG Ala Lys	CAA CAT Gln His 270	r TTT s Phe	GTT Val	GAT Asp	GCC Ala 275	CTC Leu	CTC Leu	ACA Thr	871
TTG CAA GAO Leu Gln Asp 280	AAA TAT Lys Tyr	GAC CTT Asp Leu 285	AGT GAN	A GAC u Asp	ACC Thr	ATC Ile 290	ATT Ile	GGT Gly	CTC Leu	CTT Leu	919
TGG GAT ATO Trp Asp Met 295	G ATC ACA	GCA GGG Ala Gly 300	ATG GAG	C ACA p Thr	ACT Thr 305	GCA Ala	ATT Ile	TCA Ser	GTT Val	GAG Glu 310	967
TGG GCC ATO	G GCT GAC : Ala Glu 319	ı Leu Ile	AGA AA Arg As	C CCA n Pro 320	AGG Arg	GTG Val	CAA Gln	CAA Gln	AAG Lys 325	GTC Val	1015
CAA GAG GAG Gln Glu Glu	G CTA GAC 1 Leu Ası 330	AGG GTA Arg Val	ATT GG	y Leu	GAA Glu	AGG Arg	GTG Val	ATG Met 340	ACT Thr	GAA Glu	1063
GCA GAC TTO Ala Asp Pho 34:	e Ser Ası	CTC CCT	TAC CT. Tyr Le	A CAA u Gln	TGT Cys	GTG Val	ACC Thr 355	AAA Lys	GAA Glu	GCA Ala	1111
ATG AGG CT Met Arg Le	r CAC CCA u His Pro	A CCA ACC Pro Thr 365	Pro Le	A ATG u Met	CTC Leu	CCA Pro 370	CAC His	CGT Arg	GCC Ala	AAT Asn	1159
GCC AAT GT Ala Asn Va 375	C AAA GT l Lys Val	r GGA GGC L Gly Gly 380	TAT GA Tyr As	C ATT p Ile	CCC Pro 385	AAA Lys	GGG Gly	TCC Ser	AAT Asn	GTG Val 390	1207
CAT GTG AA His Val As	T GTG TG n Val Trj 39	o Ala Val	GCC CG . Ala Ar	C GAC g Asp 400	CCG Pro	GCC Ala	GTG Val	TGG Trp	AAG Lys 405	GAT Asp	1255
CCA TTG GA Pro Leu Gl	G TTC CG. u Phe Ar	A CCC GAA	AGG TT	C CTT e Leu	GAG Glu	GAG Glu	GAT Asp	GTA Val	GAC Asp	ATG Met	1303

PCT/US98/20807

-78-

			410					415	, 0				420			
AAG Lys	GGC Gly	CAT His 425	GAC Asp	TTT Phe	AGG Arg	CTA Leu	CTT Leu 430	CCA Pro	TTC Phe	GGG Gly	TCG Ser	GGT Gly 435	CGA Arg	CGA Arg	GTA Val	1351
TGC Cys	CCG Pro 440	GGT Gly	GCC Ala	CAA Gl.n	CTT Leu	GGT Gly 445	ATC Ile	AAC Asn	TTG Leu	GCA Ala	GCA Ala 450	TCC Ser	ATG Met	TTG Leu	GGC Gly	1399
CAC His 455	CTC Leu	TTG Leu	CAC His	CAT His	TTC Phe 460	TGT Cys	TGG Trp	ACC Thr	CCA Pro	CCT Pro 465	GAA Glu	GGA Gly	ATG Met	AAG Lys	CCT Pro 470	1447
GAG Glu	GAA Glu	ATT Ile	GAC Asp	ATG Met 475	GGA Gly	GAG Glu	AAT Asn	CCA Pro	GGG Gly 480	CTA Leu	GTC Val	ACA Thr	TAC Tyr	ATG Met 485	AGG Arg	1495
ACT	CCA Pro	ATA Ile	CAA Gln 490	GCT Ala	GTG Val	GTT Val	TCT Ser	CCT Pro 495	AGG Arg	CTC Leu	CCC Pro	TCA Ser	CAT His 500	TTA Leu	TAC Tyr	1543
	CGT Arg						TAA'	TCTT	rct '	rttc'	TTTC	CC T	rgga	CTAC	T	1594
CTT	TGTT	GCA '	TTAA	GAAA.	AA T	GCCT'	rgtg	G CA	CTAC	TTTT	ATC	TTTG	TGT '	TAT	GTAACT	1654
ACA	TATG	AAA '	TCAC	AATT	TA A	GGAA(CTAA	g gai	AAAA	CTCA	TTG	CGAG	GGT			1704
(2)	INF	ORMA	TION	FOR	SEQ	ID 1	NO:1	8:								
		(i)	(A (B) LE	NGTH PE :	RACT : 50 amin GY:	9 am o ac		: acid	s						
	(ii)	MOLE													
	(xi)	SEQU	ĖNCE	DES	CRIP	TION	: SE	Q ID	ΝΟ:	18:					
Met 1		Leu	. Leu	Leu 5		Ile	Pro	Ile	Ser 10		Val	Thr	Leu	. Trp	Leu	
Gly	Tyr	Thr	Leu 20		Gln	Arg	Leu	Arg 25		Lys	Leu	Pro	Pro 30		Pro	
Arg	Pro	Trp 35		Val	Val	Gly	Asn 40		Tyr	. Asp	Ile	Lys 45		Val	Arg	
Phe	Arg		Phe	Ala	Glu	Trp 55		Gln	Ser	Tyr	Gly 60) Ile	: Ile	e Ser	
Va] 65	_	Phe	e Gly	/ Ser	Thr 70		Asr	ı Val	. Ile	val 75		· Asr	. Ser	Glu	Leu 80	
Ala	a Lys	Gli	ı Val	. Leu	Lys	Glu	His	a Asp	Glr	Leu	ı Lev	ı Ala	Asp	Arg	His	

-79-

Arg	Ser	Arg	Ser 100	Ala	Ala	Lys	Phe	Ser 105	Arg	Asp	Gly	Lys	Asp 110	Leu	Ile
Trp	Ala	Asp 115	Tyr	Gly	Pro	His	Tyr 120	Val	Lys	Val	Arg	Lys 125	Val	Cys	Thr
Leu	Glu 130	Leu	Plie	Ser	Pro	Lys 135	Arg	Leu	Glu	Ala	Leu 140	Arg	Pro	Ile	Arg
Glu 145	Asp	Glu	Val	Thr	Ser 150	Met	Val	Asp	Ser	Val 1.55	Tyr	Asn	His	Cys	Thr 150
Ser	Thr	Glu	Asn	Leu 165	Gly	Lys	Gly	Ile	Leu 170	Leu	Arg	Lys	His	Leu 175	Gly
Val	Val	Ala	Phe 180	Asn	Asn	Ile	Thr	Arg 185	Leu	Ala	Phe	Gly	Lys 190	Arg	Phe
Val	Asn	Ser 195	Glu	Gly	Val	Met	Asp 200	Glu	Gln	Gly	Val	Glu 205	Phe	Lys	Ala
Ile	Val 210	Glu	Asn	Gly	Leu	Lys 215	Leu	Gly	Ala	Ser	Leu 220	Ala	Met	Ala	Glu
His 225	Ile	Pro	Trp	Leu	Arg 230	Trp	Met	Phe	Pro	Leu 235	Glu	Glu	Gly	Ala	Phe 240
Ala	Lys	His	Gly	Ala 245	Arg	Arg	Asp	Arg	Leu 250	Thr	Arg	Ala	Ile	Met 255	Ala
Glu	His	Thr	Glu 260	Ala	Arg	Lys	Lys	Ser 265	Gly	Gly	Ala	Lys	Gln 270	His	Phe
Val	Asp	Ala 275	Leu	Leu	Thr	Leu	Gln 280	Asp	Lys	Tyr	Asp	Leu 285	Ser	Glu	Asp
Thr	Ile 290	Ile	Gly	Leu	Leu	Trp 295	Asp	Met	Ile	Thr	Ala 300	Gly	Met	Asp	Thr
Thr 305	Ala	Ile	Ser	Val	Glu 310	Trp	Ala	Met	Ala	Glu 315	Leu	Ile	Arg	Asn	Pro 320
Arg	Val	Gln	Gln	Lys 325	Val	Gln	Glu	Glu	Leu 330	Asp	Arg	Val	Ile	Gly 335	Leu
Glu	Arg	Val	Met 340	Thr	Glu	Ala	Asp	Phe 345	Ser	Asn	Leu	Pro	Tyr 350	Leu	Gln
Cys	Val	Thr 355	_	Glu	Ala	Met	Arg 360		His	Pro	Pro	Thr 365	Pro	Leu	Met
Leu	Pro 370	His	Arg	Ala	Asn	Ala 375	Asn	Val	Lys	Val	Gly 380	Gly	Tyr	Asp	Ile
Pro 385	-	Gly	Ser	Asn	Val 390	His	Val	Asn	Val	Trp 395	Ala	Val	Ala	Arg	Asp 400
Pro	Ala	Val	Trp	Lys 405		Pro	Leu	Glu	Phe 410		Pro	Glu	Arg	Phe 415	Leu

WO 99/19493 PCT/US98/20807

-80-

Glu Glu Asp Val Asp Met Lys Gly His Asp Phe Arg Leu Leu Pro Phe
420 425 430

Gly Ser Gly Arg Arg Val Cys Pro Gly Ala Gln Leu Gly Ile Asn Leu
435 440 445

Ala Ala Ser Met Leu Gly His Leu Leu His His Phe Cys Trp Thr Pro 450 455 460

Pro Glu Gly Met Lys Pro Glu Glu Ile Asp Met Gly Glu Asn Pro Gly 465 470 475 480

Leu Val Thr Tyr Met Arg Thr Pro Ile Gln Ala Val Val Ser Pro Arg
485 490 495

Leu Pro Ser His Leu Tyr Lys Arg Val Pro Ala Glu Ile 500 505

- (2) INFORMATION FOR SEQ ID NO:19:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TGTCTAACTC CTTCCTTTTC

20

- (2) INFORMATION FOR SEQ ID NO:20:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Phe Leu Pro Phe Gly Xaa Gly Xaa Arg Xaa Cys Xaa Gly

(2) INFORMATION FOR SEQ ID NO:21:

-81-

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Phe Xaa Xaa Gly Xaa Xaa Xaa Cys Xaa Gly

- (2) INFORMATION FOR SEQ ID NO:22:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Xaa Cys Xaa Gly

- (2) INFORMATION FOR SEQ ID NO:23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Pro Glu Glu Phe Xaa Pro Glu Arg Phe 1 5

10

THAT WHICH IS CLAIMED IS:

- 1. An isolated DNA molecule comprising a sequence selected from the group consisting of:
 - a) SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, and SEO ID NO:17;
 - b) DNA sequences which encode an enzyme having a sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6; SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, and SEQ ID NO:18;
 - c) DNA sequences which have at least about 90% sequence identity to the DNA of (a) or (b) above and which encode a cytochrome P450 enzyme; and
 - d) DNA sequences which differ from the DNA of (a) or (c) above due to the degeneracy of the genetic code.
 - 2. A peptide encoded by a DNA sequence of claim 1.
- 3. A cytochrome p450 enzyme having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO6:, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, and SEQ ID NO:18.
- 4. An isolated DNA molecule comprising a sequence selected from the group consisting of:
 - a) SEQ ID NO:1;
- b) DNA sequences which encode an enzyme having SEQ ID NO:2,;
 - c) DNA sequences which have at least about 90% sequence identity to the DNA of (a) or (b) above and which encode a cytochrome P450 enzyme; and

- d) DNA sequences which differ from the DNA of (a) or (c) above due to the degeneracy of the genetic code.
 - 5. A peptide encoded by a DNA sequence of claim 4.
 - 6. A cytochrome p450 peptide having SEQ ID NO:2.
 - 7. A DNA construct comprising an expression cassette, which construct comprising in the 5' to 3' direction, a promoter operable in a plant cell and a DNA segment according to claim 1 positioned downstream from said promoter and operatively associated therewith.
 - 8. A DNA construct according to claim 7, wherein said promoter is constitutively active in plant cells.
 - 9. A DNA construct according to claim 7, wherein said promoter is the 35S promoter from Cauliflower Mosaic virus.
 - 10. A DNA construct according to claim 7, said construct further comprising a plasmid.
 - 11. A DNA construct according to claim 7 carried by a plant transformation vector.
 - 12. A DNA construct according to claim 7 carried by an Agrobacterium tumefaciens plant transformation vector.
 - 13. A plant cell containing a DNA construct according to claim 7.
 - 14. A transgenic plant comprising plant cells according to claim 13.

- 15. A transgenic plant according to claim 14, wherein said plant is a monocot.
- 16. A transgenic plant according to claim 14, wherein said plant is a dicot.
- 17. A DNA construct comprising an expression cassette, which construct comprising in the 5' to 3' direction, a promoter operable in a plant cell, and a DNA segment encoding a peptide of SEQ ID NO:2 positioned downstream from said promoter and operatively associated therewith.
- 18. A DNA construct according to claim 17, wherein said promoter is constitutively active in plant cells.
- 19. A DNA construct according to claim 17, wherein said promoter is the35S promoter from Cauliflower Mosaic virus.
- 20. A DNA construct according to claim 17, said construct further comprising a plasmid.
- 21. A DNA construct according to claim 17 carried by a plant transformation vector.
- 22. A DNA construct according to claim 17 carried by an Agrobacterium tumefaciens plant transformation vector.
 - 23. A plant cell containing a DNA construct according to claim 17.
 - 24. A transgenic plant comprising plant cells according to claim 23.

- 25. A transgenic plant according to claim 24, wherein said plant is a monocot.
- 26. A transgenic plant according to claim 24, wherein said plant is a dicot.
- 27. A method of making a transgenic plant cell having an increased ability to metabolize phenylurea compounds compared to an untransformed plant cell, said method comprising:
 - a) providing a plant cell;
 - b) transforming said plant cell with an exogenous DNA construct comprising, in the 5' to 3' direction, a promoter operable in a plant cell and a DNA sequence encoding a peptide of SEQ ID NO:2, said DNA sequence operably linked to said promoter.
- 28. A method according to claim 27, wherein said plant cell is from a member of the Solanacae family.
- 29. A method according to claim 27, wherein said promoter is the 35S promoter from Cauliflower Mosaic virus.
- 30. A method according to claim 27, wherein said transforming step is carried out by bombarding said plant cell with microparticles carrying said DNA construct.
- 31. A method according to claim 27 wherein said transforming step is carried out by infecting said plant cell with an *Agrobacterium tumefaciens* containing a Ti plasmid carrying said DNA construct.
- 32. A method according to claim 27, further comprising regenerating a plant from said transformed plant cell.

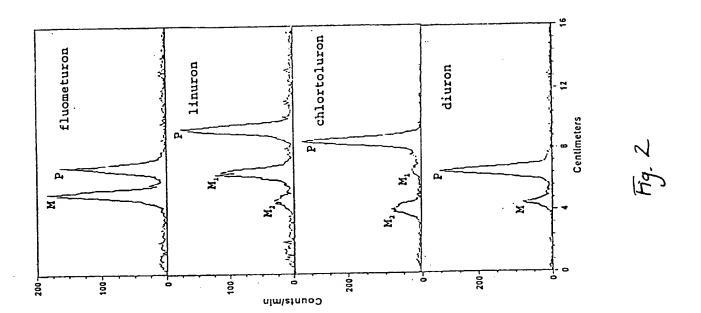
PCT/US98/20807

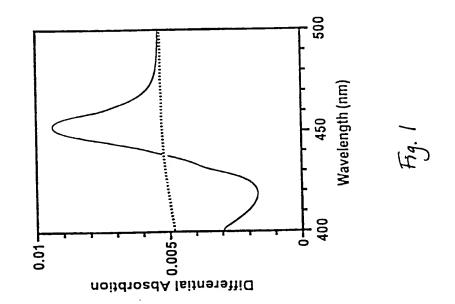
- 33. A transformed plant produced by the method of claim 32.
- 34. Seed or progeny of a plant according to claim 33, which seed or progeny has inherited said DNA sequence encoding a peptide of SEQ ID NO:2.
- 35. A transformed plant produced by the method of claim 32, which plant has increased resistance to phenylurea herbicides compared to wild-type plants of the same species.
- 36. A transgenic plant having an increased ability to metabolize phenylurea compounds compared to an untransformed plant cell, said transgenic plant comprising transgenic plant cells containing an exogenous DNA construct comprising, in the 5' to 3' direction, a promoter operable in said plant cell, said promoter operably linked to a DNA sequence encoding a peptide of SEQ ID NO:2.
- 37. A transgenic plant according to claim 36, wherein said promoter is the 35S promoter from Cauliflower Mosaic virus.
- 38. A transgenic plant according to claim 36, wherein said plant is a dicot.
- 39. A transgenic plant according to claim 36, wherein said plant is a monocot.
- 40. A transgenic plant according to claim 36, wherein said plant is a member of the family Solanacae.
- 41. A transgenic plant according to claim 36, which plant is selected from the group consisting of tobacco, potato, tomato, corn, rice, cotton, soybean,

rape, wheat, oats, barley, rye and rice.

- 42. Progeny or seed of a plant according to claim 36, wherein said seed or progeny has inherited said DNA sequence encoding a peptide of SEQ ID NO:2.
- 43. A transformed plant according to claim 36, which plant has increased resistance to phenylurea herbicides compared to wild-type plants of the same species.
- 44. A crop comprising a plurality of plants according to claim 36 planted in an agricultural field.
- 45. A method of using a phenylurea herbicide as a post-emergence, herbicide, comprising:
 - a) planting a crop according to claim 44;
 - b) applying to said crop a phenylurea herbicide.
- 46. A method according to claim 45, wherein said crop is selected from the group consisting of turfgrass, tobacco, potato, tomato, corn, rice, cotton, soybean, rape, wheat, oats, barley, rye and rice.
- 47. A method according to claim 45, wherein said herbicide is selected from the group consisting of fluometuron, linuron, chlortoluron and diuron.

1/3





PCT

LD INTELLECTUAL PROPERTY ORGANIZATIO. International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶:
C12N 15/53, 15/82, 9/02, 5/00, A01H
5/00, 5/10

A3

(11) International Publication Number:

WO 99/19493

1 ...

(43) International Publication Date:

22 April 1999 (22.04.99)

(21) International Application Number:

PCT/US98/20807

(22) International Filing Date:

5 October 1998 (05.10.98)

(30) Priority Data:

08/948,564

10 October 1997 (10.10.97) US

10 October 1997 (10.10.97)

(71) Applicant (for all designated States except US): NORTH CAROLINA STATE UNIVERSITY [US/US]; 103 Bynum Hall, CB #7003, Raleigh, NC 27695-7003 (US).

(72) Inventors; and

- (75) Inventors/Applicants (for US only): SIMINSZKY, Balazs [HU/US]; 708 F Chapel Hill Drive, Raleigh, NC 27606 (US). DEWEY, Ralph, E. [US/US]; 9432 Cartersville Court, Raleigh, NC 27613 (US). CORBIN, Frederick, T. [US/US]; 4508 Lead Mine Road, Raleigh, NC 27612 (US).
- (74) Agents: BENNETT, Virginia, C. et al.; Myers, Bigel, Sibley & Sajovec, P.A., P.O. Box 37428, Raleigh, NC 27627 (US).

(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CiN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model) GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC. LK, LR, LS, LT, LU, LV, MP, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims

and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
1 July 1999 (01.07.99)

(54) Title: CYTOCHROME P-450 CONSTRUCTS AND METHODS OF PRODUCING HERBICIDE-RESISTANT TRANSGENIC PLANTS

(57) Abstract

DNA sequence encoding novel cytochrome P-450 molecules are provided. The use of DNA constructs containing such molecules to transform plants is described, as are transgenic plants exhibiting increased resistance to phenylurea herbicides. Methods of using such DNA constructs and transformed plants are provided.

^{*(}Referred to in PCT Gazette No. 31/1999, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL.	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	ТJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
ВJ	Benin	1E	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CÜ	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
ÐE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	1.R	Liberia	SG	Singapore		

WO 99/19493 PCT/US98/20807

CYTOCHROME P-450 CONSTRUCTS AND METHODS OF PRODUCING HERBICIDE-RESISTANT TRANSGENIC PLANTS

Field of the Invention

The present invention relates to DNA encoding nevel cytochrome P-450 molecules, and the transformation of cells with such DNA. These DNA sequences may be used in methods of producing plants with an altered ability to metabolize chemical compounds, such as phenylurea herbicides.

Background of the Invention

Cytochrome P-450 (P-450) monooxygenases are ubiquitous hemoproteins present in microorganisms, plants and animals. Comprised of a large and diverse group of isozymes, P-450s mediate a great array of oxidative reactions using a wide range of compounds as substrates, and including biosynthetic processes such as phenylpropanoid, fatty acid, and terpenoid biosynthesis; metabolism of natural products; and detoxification of foreign substances (xenobiotics). See e.g., Schuler, Crit. Rev. Plant Sci. 15:235-284 (1996). In a typical P-450 catalyzed reaction, one atom of molecular oxygen (O₂) is incorporated into the substrate, and the other atom is reduced to water by NADPH. For most eucaryotic P-450s, NADPH:cytochrome P-450 reductase, a membrane-bound flavoprotein, transfers the necessary two electrons from NADPH to the P-450 (Bolwell et al, Phytochemistry 37: 1491-1506 (1994)).

Frear et al. (Phytochemistry 8:2157-2169 (1969)) demonstrated the metabolism of monuron by a mixed-function oxidase located in a microsomal fraction of cotton seedlings. Further evidence has accumulated supporting the involvement of P-450s in the metabolism and detoxification of numerous herbicides representing several distinct classes of compounds (reviewed in Bolwell et al., 1994; Schuler, 1996). Differential herbicide metabolizing P-450 activities are believed to represent one of the mechanisms that enables certain crop species to be more tolerant of a particular herbicide than other crop or weedy species.

5

10

15

20

10

15

20

25

Summary of the Invention

A first aspect of the present invention is an isolated DNA molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, or SEQ ID NO:17; or DNA sequences which encode an enzyme of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18; or DNA sequences which have at least about 90% sequence identity to the above DNA and which encode a cytochrome P450 enzyme; and DNA sequences which differ from the above DNA due to the degeneracy of the genetic code.

A further aspect of the present invention is a cytochrome p450 enzyme having an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6; SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18.

A further aspect of the present invention is an isolated DNA molecule comprising SEQ ID NO:1; DNA sequences which encode an enzyme of SEQ ID NO:2,; DNA sequences which have at least about 90% sequence identity to the above DNA and which encode a cytochrome P450 enzyme; and DNA sequences which differ from the above DNA due to the degeneracy of the genetic code.

A further aspect of the present invention is a cytochrome p450 peptide of SEO ID NO:2.

A further aspect of the present invention is a DNA construct comprising a promoter operable in a plant cell and a DNA segment encoding a peptide of SEQ ID NO:2 downstream from and operatively associated with the promoter.

A further aspect of the present invention is a method of making a transgenic plant cell having an increased ability to metabolize phenylurea compounds compared to an untransformed plant cell. The plant cell is transformed with an exogenous DNA construct comprising a promoter operable in a plant cell and a DNA sequence encoding a peptide of SEQ ID NO:2.

30 Transformed plants, seed and progeny of such plants are also aspects of the

present invention.

A further aspect of the present invention is a transgenic plant having an increased ability to metabolize phenylurea compounds. Such transgenic plants contain exogenous DNA encoding a peptide of SEQ ID NO:2.

5

10

15

20

25

30

Brief Description of the Drawings

Figure 1 depicts dithionite-reduced carbon monoxide, difference spectra, where the solid line represents microsomes isolated from yeast transformed with CYP71A10, and the dotted line shows the difference spectra from yeast transformed with control vector V-60. Microsomal protein concentration was 1 mg/ml.

Figure 2 shows thin-layer chromatograms of [14 C]-radiolabeled fluometuron, linuron, chlortoluron, and diuron and their respective metabolites after incubation of the radiolabeled herbicides with yeast microsomes containing the CYP71A10 protein. Initial substrate concentrations for fluometuron, linuron, chlortoluron and diuron were 5.2, 6.5, 4.0, and 3.7 μ M, respectively. P = parent compound; M = metabolite.

Figure 3 shows the chemical structures of fluometuron, linuron, chlortoluron and diuron, and their previously characterized metabolites. The linuron and chlortoluron metabolites are designated major or minor depending on their predicted relative abundance in assays using yeast microsomes containing the soybean CYP71A10 protein.

Figure 4 shows thin-layer chromatograms using [14 C]-radiolabeled linuron in various control reactions. The complete reaction mixture (COMPLETE) contained 3.2 μ M linuron, 0.75 mM NADPH and 2.5 mg/ml microsomal protein isolated from CYP71A10-transformed yeast in 50 mM phosphate buffer (pH 7.1). Other reactions varied from COMPLETE by the addition of carbon monoxide (+CO), the omission of NADPH (NO NADPH), or the use of yeast microsomes isolated from cells expressing the control vector (V-60). P = parent compound; M = metabolite.

10

15

20

25

30

Figure 5A shows tobacco line 25/2 plants (transformed with soybean CYP71A10) grown on media containing no herbicide.

Figure 5B shows control tobacco plants (transformed with vector pBI121) grown on media containing 0.5 µM linuron.

Figure 5C shows tobacco line 25/2 (transformed with soybean CYP71A10) individuals grown on media containing 0.5 μ M linuron.

Figure 5D shows tobacco line 25/2 (transformed with soybean CYP71A10) individuals grown on media containing 2.5 μ M linuron.

Figure 5E shows control tobacco plants (transformed with vector pBI121) grown on media containing 1.0 µM chlortoluron.

Figure 5F shows tobacco line 25/2 (transformed with soybean CYP71A10) individuals grown on media containing 1.0 μ M chlortoluron.

Detailed Description of the Invention

1. Overview of the present research:

The present inventors utilized a strategy based on the random isolation and screening of soybean cDNAs encoding cytochrome P-450 (P-450) isozymes to identify P-450 isozymes involved in herbicide metabolism. Eight full-length and one near full-length P-450 cDNAs representing eight distinct P-450 families were isolated using polymerase chain reaction (PCR)-based technologies (SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15 and 17). Five of these soybean P-450 cDNAs were successfully overexpressed in yeast, and microsomal fractions generated from these strains were tested for their potential to mediate the metabolism of ten herbicides and one insecticide. *In vitro* enzyme assays showed that the gene product of one heterologously expressed P-450 cDNA (CYP71A10) (SEQ ID NO:1) specifically mediated the metabolism of phenylurea herbicides, converting four herbicides of this class (fluometuron, linuron, chlortoluron, and diuron) into more polar metabolites. Analyses of the metabolites indicate that the CYP71A10 encoded enzyme functions primarily as an N-demethylase with regard to

10

15

20

25

30

fluometuron, linuron and diuron, and as a ring-methyl hydroxylase when chlortoluron is the substrate. *In vivo* assays using excised leaves demonstrated that all four herbicides were more readily metabolized in CYP71A10-transformed tobacco in comparison to control plants.

Shiota et al. reported that fused constructs derived from the rat CYP1A1 and yeast NADPH-cytochrome P-450 oxidoreductase cDNAs conferred chlortoluron resistance in tobacco by enhancing herbicide metabolism (Shiota et al., Plant Physiol. 106:17-23 (1994)). In another study, a chloroplast-targeted, bacterial CYP105A1 expressed in tobacco catalyzed the toxification of R7402, a sulfonylurea pro-herbicide (O'Keefe et al., Plant Physiol. 105:473-482 (1994)). The cloning and heterologous expression of an endogenous plant P-450 gene that is potentially involved in herbicide metabolism was reported by Pierrel et al., Eur. J. Biochem. 224:835-844 (1994), where a trans-cinnamic acid hydroxylase cDNA (CYP73A1) isolated from artichoke and expressed in yeast catalyzed the ring-methyl hydroxylation of chlortoluron. In vivo experiments with artichoke tubers, however, demonstrated that the ring-methyl hydroxy metabolite represented only a minor portion of the metabolites produced and that the major metabolite was demethylated chlortoluron (Pierrel et al., 1994). This together with the observation that the turnover number of the heterologously expressed enzyme was very low (0.014/ min), suggested that CYP73A1 plays a minimal role in chlortoluron metabolism in vivo. US Patent No. 5,349,127 to Dean et al. discloses the use of DNA encoding certain P-450 enzymes, isolated from Streptomyces griseolus, to produce transformed plants with increased metabolism of certain compounds. (All US patents referred to herein are intended to be incorporated herein in their entirety.)

Although the role of P-450 enzymes in catalyzing the metabolism of a variety of herbicides has been documented, little progress has been made in the identification of the endogenous plant P-450s that are responsible for degrading these compounds. Protein purification of specific isozymes involved in the metabolism of a specific herbicide has been hindered by the instability of the

10

15

20

25

30

enzymes, their low concentrations in most plant tissues, and difficulties in the reconstitution of active complexes from solubilized components. Furthermore, any given plant tissue may possess dozens, if not hundreds, of unique P-450 isozymes, complicating the purification to homogeneity of a particular isozyme. Because plants have only been exposed to phenylurea herbicides during the past few decades, it is unlikely that enzymes have evolved solely for the purposed of metabolizing this class of xenobiotics.

2. Use of CYP71A10 to produce phenylurea-resistant plants:

The present invention provides materials and methods useful in producing transgenic plant cells and plants with increased resistance to phenylurea herbicides. Increased herbicide resistance, as used herein, refers to the ability of a plant to withstand levels of an herbicide that have a negative impact on wildtype (untransformed) plants of the same species and/or variety. Resistance, as used herein, does not necessarily mean that the resistant plant is completely unaffected by exposure to the herbicide; rather, resistant plants suffer less extensive or less severe damage than comparable wild-type plants. Methods of assessing the extent and/or severity of herbicide impact will vary depending on the particular plant and the particular herbicide being tested; such assessment methods will be apparent to those skilled in the art. The negative effects of a herbicide may be evidenced by the complete arrest of plant growth, or by an inhibition in the rate or amount of growth. Additionally, methods of the present invention may be used to decrease herbicide residues in plants, even where the amounts of herbicides present in the plant do not cause an appreciable negative effect on the plant as a whole.

Increased resistance to a herbicide can be due to an increased ability to metabolize a herbicide to less harmful metabolites. Accordingly, plants of the present invention which exhibit increased resistance to a herbicide may also be described as having an increased ability to metabolize the starting herbicidal compound, where the metabolites are less harmful to the plant than the starting

در الميارة در الميارة

Ager

- 1-75

5

10

15

20

25

30

compound.

In the examples provided herein, yeast microsomes and transgenic tobacco plants expressing the CYP71A10 peptide (SEQ ID NO:2) and exposed to various phenylurea herbicides produced the same degradation products that have previously been observed when these same compounds have been incubated with metabolically active plant microsomes. These results indicate that the CYP71A10 peptide plays a role in the effective metabolism of phenylurea herbicides.

The present examples demonstrate that the overexpression of a CYP71A10 peptide of SEQ ID NO:2 in tobacco enhanced the plant's capacity to metabolize all four phenylurea herbicides tested, and that appreciable levels of tolerance were conferred to linuron and chlortoluron. Fluometuron was the most actively metabolized compound in both the yeast and transgenic plant systems, yet the enhancement in tolerance to this herbicide at the whole plant level was not as great as for linuron and chlortoluron. While not wishing to be held to a single theory, the present inventors surmise that the lack of correlation between the rate of herbicide metabolism and herbicide tolerance may be explained by the differential toxicities of the various phenylurea derivatives produced in the Consistent with this hypothesis are the CYP71A10-transformed tobacco. previous observations that N-demethyl derivatives of fluometuron, diuron and chlortoluron are only moderately less toxic than their parent compounds (Rubin and Eshel, Weed Sci. 19:592-594 (1971); Dalton et al., Weeds 14:31-33 (1966); Ryan and Owen, Proc. Brit. Crop Prot. Conf. Weeds 1:317-324 (1982)). contrast, linuron is a 10-fold greater inhibitor of the Hill-reaction than demethyl linuron (Suzuki and Casida, J. Agric. Food Chem. 29:1027-1033 (1981)), and the hydroxylated and the didemethlayed derivatives of chlortoluron are considered to be nonherbicidal (Ryan and Owen, 1982).

The present inventors found that the relative rates of herbicide metabolism in leaves of CYP71A10-transformed tobacco and in yeast microsomes assayed in vitro were similar (see Tables 4 and 5). With the exception of the transgenic

10

15

20

25

30

plant leaves showing a somewhat greater metabolic activity against chlortoluron than was apparent in the yeast microsomal assays, both systems followed the general order of metabolism of fluometuron \geq linuron > chlortoluron > diuron. These results indicate that expression of a test plant P-450 in yeast and quantification of the metabolism of a test compound using yeast microsomes, is a suitable system for screening plant P-450s for their metabolic function, and for their potential usefulness in the production of transgenic plants with altered metabolism of chemical compounds such as herbicides and insecticides.

The present inventors have shown that the random isolation of P-450 cDNAs with subsequent heterologous expression in yeast is an effective strategy to characterize cDNAs whose product is capable of affecting the metabolism of a test compound. This approach is useful in characterizing the substrates (both natural and artificial) affected by a P-450, in determining the function of P-450 genes whose catalytic activities remain unclear, and in screening P-450s for the ability to increase or decrease the metabolism of a test compound. A particularly useful aspect of this method is the ability to screen isolated P-450s for their effects on the metabolism by plants of herbicides, insecticides, or other chemical compounds. Increased metabolism may result in enhanced resistance to the effects of a compound (where the metabolites are less harmful than the starting compound), or in increased sensitivity to the effects of a compound (where one or more metabolites are more toxic than the starting compound; see O'Keefe et al., 1994).

3. DNA Constructs:

Those familiar with recombinant DNA methods available in the art will recognize that one can employ a cDNA molecule (or a chromosomal gene or genomic sequence) encoding a P-450 peptide, joined in the sense orientation with appropriate operably linked regulatory sequences, to construct transgenic cells and plants. (Those of skill in the art will also recognize that appropriate regulatory sequences for expression of genes in the sense orientation include any

10

15

20

25

30

one of the known eukaryotic translation start sequences, in addition to the promoter and polyadenylation/transcription termination sequences described herein). Appropriate selection of the encoded P-450 peptide will provide transformed plants characterized by altered (enhanced or retarded) metabolism of phenylurea compounds.

DNA constructs, or "transcription cassettes," of the present invention include, 5' to 3' in the direction of transcription, a promoter as discussed herein, a DNA sequence as discussed herein operatively associated with the promoter, and, optionally, a termination sequence including stop signal for RNA polymerase and a polyadenylation signal for polyadenylase. All of these regulatory regions should be capable of operating in the cells of the tissue to be transformed. Any suitable termination signal may be employed in carrying out the present invention, examples thereof including, but not limited to, the nopaline synthase (nos) terminator, the octapine synthase (ocs) terminator, the CaMV terminator, or native termination signals derived from the same gene as the transcriptional initiation region or derived from a different gene. See, e.g., Rezian et al. (1988) supra, and Rodermel et al. (1988), supra.

The term "operatively associated," as used herein, refers to DNA sequences on a single DNA molecule which are associated so that the function of one is affected by the other. Thus, a promoter is operatively associated with a DNA when it is capable of affecting the transcription of that DNA (i.e., the DNA is under the transcriptional control of the promoter). The promoter is said to be "upstream" from the DNA, which is in turn said to be "downstream" from the promoter.

The transcription cassette may be provided in a DNA construct which also has at least one replication system. For convenience, it is common to have a replication system functional in Escherichia coli, such as ColE1, pSC101, pACYC184, or the like. In this manner, at each stage after each manipulation, the resulting construct may be cloned, sequenced, and the correctness of the manipulation determined. In addition, or in place of the E. coli replication

15

20

25

30

system, a broad host range replication system may be employed, such as the replication systems of the P-1 incompatibility plasmids, e.g., pRK290. In addition to the replication system, there will frequently be at least one marker present, which may be useful in one or more hosts, or different markers for individual hosts. That is, one marker may be employed for selection in a prokaryotic host, while another marker may be employed for selection in a eukaryotic host, particularly the plant host. The markers may be protection against a biocide, such as antibiotics, toxins, heavy metals, or the like; may provide complementation, by imparting prototrophy to an auxotrophic host; or may provide a visible phenotype through the production of a novel compound in the plant.

The various fragments comprising the various constructs, transcription cassettes, markers, and the like may be introduced consecutively by restriction enzyme cleavage of an appropriate replication system, and insertion of the particular construct or fragment into the available site. After ligation and cloning the DNA construct may be isolated for further manipulation. All of these techniques are amply exemplified in the literature as exemplified by J. Sambrook et al., Molecular Cloning, A Laboratory Manual (2d Ed. 1989)(Cold Spring Harbor Laboratory).

Vectors which may be used to transform plant tissue with nucleic acid constructs of the present invention include both Agrobacterium vectors and ballistic vectors, as well as vectors suitable for DNA-mediated transformation.

4. Promoters:

The term 'promoter' refers to a region of a DNA sequence that incorporates the necessary signals for the efficient expression of a coding sequence. This may include sequences to which an RNA polymerase binds but is not limited to such sequences and may include regions to which other regulatory proteins bind together with regions involved in the control of protein translation and may include coding sequences.

10

15

Promoters employed in carrying out the present invention may be constitutively active promoters. Numerous constitutively active promoters which are operable in plants are available. A preferred example is the Cauliflower Mosaic Virus (CaMV) 35S promoter which is expressed constitutively in most plant tissues. Use of the CaMV promoter for expression of recombinant genes in tobacco roots has been well described (Lam et al., "Site-Specific Mutations Alter In Vitro Factor Binding and Change Promoter Expression Pattern in Transgenic Plants", Proc. Nat. Acad. Sci. USA 86, pp. 7890-94 (1989); Poulsen et al. "Dissection of 5' Upstream Sequences for Selective Expression of the Nicotiana plumbaginifolia rbcS-8B Gene", Mol. Gen. Genet. 214, pp. 16-23 (1988)). In the alternative, the promoter may be a tissue-specific promoter or a promoter that is expressed temporally or developmentally. See, e.g., US Patent No. 5,459,252 to Conkling et al.; Yamamoto et al., The Plant Cell, 3:371 (1991). In methods of transforming plants to alter the effects of herbicides or to decrease residual amounts of herbicides or pesticides in plants, selection of a suitable promoter will vary depending on the plant species, the specific chemical compound used as a herbicide or pesticide, and the time and method of applying the chemical compound to the plant or plant crop, as will be apparent to those skilled in the art.

20

25

30

5. Selectable Markers:

The recombinant DNA molecules and vectors used to produce the transformed cells and plants of this invention may further comprise a dominant selectable marker gene. Suitable dominant selectable markers include, inter alia, antibiotic resistance genes encoding neomycin phosphotransferase (NPTII), hygromycin phosphotransferase (HPT), and chloramphenicol acetyltransferase (CAT). Another well-known dominant selectable marker suitable is a mutant dihydrofolate reductase gene that encodes methotrexate-resistant dihydrofolate reductase. DNA vectors containing suitable antibiotic resistance genes, and the corresponding antibiotics, are commercially available. Transformed cells are

selected out of the surrounding population of non-transformed cells by placing the mixed population of cells into a culture medium containing an appropriate concentration of the antibiotic (or other compound normally toxic to the untransformed cells) against which the chosen dominant selectable marker gene product confers resistance. Thus, only those cells that have been transformed will survive and multiply.

A further aspect of the present invention is use of the identified P-450 coding sequences as a selectable marker gene. A DNA construct comprising a sequence encoding a P-450 known to increase resistance to a compound (such as SEQ ID NO:2) is utilized to transform cells, in accordance with methods known in the art. Those cells that subsequently exhibit resistance to the compound are indicated as transformed. Such constructs may be used to verify the success of a transformation technique or to select transformed cells of interest.

15

20

25

30

10

5

6. Sequence similarity and hybridization conditions:

Nucleic acid sequences employed in carrying out the present invention include those with sequence similarity to SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17, and encoding a protein having P-450 enzymatic activity. This definition is intended to encompass natural allelic variants and minor sequence variations in the nucleic acid sequence encoding a P-450 molecule, or minor sequence variations in the amino acid sequence of the encoded product. Thus, DNA sequences that hybridize to DNA of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17 and code for expression of a P-450 enzyme, particularly a plant P-450 enzyme, may also be employed in carrying out aspects of the present invention. The nomenclature for P-450 genes is based on amino acid sequence identity; methods of determining sequence similarity are well-known to those skilled in the art. Typically, sequences sharing >40% identity are placed in the same family, >55% identity defines members of the same subfamily, and sequences that

10

15

20

25

30

display >97% identity are assumed to represent allelic variants. Conditions which permit other DNA sequences which code for expression of a protein having P-450 enzymatic activity to hybridize to DNA of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17, or to other DNA sequences encoding the protein given as SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16 or 18 can be determined in a routine manner. For example, hybridization of such sequences may be carried out under conditions of reduced stringency or even stringent conditions (e.g., conditions represented by a wash stringency of 0.3 M NaCl, 0.03 M sodium citrate, 0.1% SDS at 60°C or even 70°C to DNA encoding the protein given as SEQ ID NO:2 herein in a standard in situ hybridization assay. See J. Sambrook et al., Molecular Cloning, A Laboratory Manual (2d Ed. 1989)(Cold Spring Harbor Laboratory)). In general, such sequences will be at least 65% similar, 75% similar, 80% similar, 85% similar, 90% similar, 93% similar, 95% similar, or even 97% or 98% similar, or more, with the sequence given herein as SEQ ID NO:1, or DNA sequences encoding proteins of SEQ ID NO:2. (Determinations of sequence similarity are made with the two sequences aligned for maximum matching; gaps in either of the two sequences being matched are allowed in maximizing matching. Gap lengths of 10 or less are preferred, gap lengths of 5 or less are more preferred, and gap lengths of 2 or less still more preferred.)

As used herein, the term 'gene' refers to a DNA sequence that incorporates (1) upstream (5') regulatory signals including a promoter, (2) a coding region specifying the product, protein or RNA of the gene, (3) downstream (3') regions including transcription termination and polyadenylation signals and (4) associated sequences required for efficient and specific expression.

The DNA sequence of the present invention may consist essentially of a sequence provided herein (SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15 or 17), or equivalent nucleotide sequences representing alleles or polymorphic variants of these genes, or coding regions thereof.

Use of the phrase "substantial sequence similarity" in the present

10

15

20

25

specification and claims means that DNA, RNA or amino acid sequences which have slight and non-consequential sequence variations from the actual sequences disclosed and claimed herein are considered to be equivalent to the sequences of the present invention. In this regard, "slight and non-consequential sequence variations" mean that "similar" sequences (i.e., the sequences that have substantial sequence similarity with the DNA, RNA, or proteins disclosed and claimed herein) will be functionally equivalent to the sequences disclosed and claimed in the present invention. Functionally equivalent sequences will function in substantially the same manner to produce substantially the same compositions as the nucleic acid and amino acid compositions disclosed and claimed herein.

DNA sequences provided herein can be transformed into a variety of host cells. A variety of suitable host cells, having desirable growth and handling properties, are readily available in the art.

Use of the phrase "isolated" or "substantially pure" in the present specification and claims as a modifier of DNA, RNA, polypeptides or proteins means that the DNA, RNA, polypeptides or proteins so designated have been separated from their *in vivo* cellular environments through the efforts of human beings.

As used herein, a "native DNA sequence" or "natural DNA sequence" means a DNA sequence which can be isolated from non-transgenic cells or tissue. Native DNA sequences are those which have not been artificially altered, such as by site-directed mutagenesis. Once native DNA sequences are identified, DNA molecules having native DNA sequences may be chemically synthesized or produced using recombinant DNA procedures as are known in the art. As used herein, a native plant DNA sequence is that which can be isolated from non-transgenic plant cells or tissue.

7. Transformed plants:

Methods of making recombinant plants of the present invention, in general, involve first providing a plant cell capable of regeneration (the plant cell

15

20

25

30

invention.

1

а

iq

ia

ic

ci

25

hi

ri.

гn.

transformed with a DNA construct comprising a transcription cassette of the present invention (as described herein) and a recombinant plant is regenerated from the transformed plant cell. As explained below, the transforming step is carried out by techniques as are known in the art, including but not limited to bombarding the plant cell with microparticles carrying the transcription cassette, infecting the cell with an Agrobacterium tumefaciens containing a Ti plasmid carrying the transcription cassette, or any other technique suitable for the production of a transgenic plant.

Numerous Agrobacterium vector systems useful in carrying out the present invention are known. For example, U.S. Patent No. 4,459,355 discloses a method for transforming susceptible plants, including dicots, with an Agrobacterium strain containing the Ti plasmid. The transformation of woody plants with an Agrobacterium vector is disclosed in U.S. Patent No. 4,795,855. Further, U.S. Patent No. 4,940,838 to Schilperoort et al. discloses a binary Agrobacterium vector (i.e., one in which the Agrobacterium contains one plasmid having the vir region of a Ti plasmid but no T region, and a second plasmid having a T region but no vir region) useful in carrying out the present

Microparticles carrying a DNA construct of the present invention, which microparticle is suitable for the ballistic transformation of a plant cell, are also useful for making transformed plants of the present invention. The microparticle is propelled into a plant cell to produce a transformed plant cell, and a plant is regenerated from the transformed plant cell. Any suitable ballistic cell transformation methodology and apparatus can be used in practicing the present invention. Exemplary apparatus and procedures are disclosed in Sanford and Wolf, U.S. Patent No. 4,945,050, and in Christou et al., U.S. Patent No. 5,015,580. When using ballistic transformation procedures, the transcription cassette may be incorporated into a plasmid capable of replicating in or integrating into the cell to be transformed. Examples of microparticles suitable

10

15

20

25

30

for use in such systems include 1 to $5 \,\mu m$ gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

Plant species may be transformed with the DNA construct of the present invention by the DNA-mediated transformation of plant cell protoplasts and subsequent regeneration of the plant from the transformed protoplasts in accordance with procedures well known in the art. Fusion of tobacco protoplasts with DNA-containing liposomes or via electroporation is known in the art. (Shillito et al., "Direct Gene Transfer to Protoplasts of Dicotyledonous and Monocotyledonous Plants by a Number of Methods, Including Electroporation", Methods in Enzymology 153, pp. 313-36 (1987)).

As used herein, transformation refers to the introduction of exogenous DNA into cells, so as to produce transgenic cells stably transformed with the exogenous DNA. Transformed plant cells are induced to regenerate intact plants through application of cell and tissue culture techniques that are well known in the art. The method of plant regeneration is chosen so as to be compatible with the method of transformation. The stable presence and the orientation of the exogenous DNA in transgenic plants can be verified by Mendelian inheritance of the DNA sequence, as revealed by standard methods of DNA analysis applied to progeny resulting from controlled crosses.

Plants of horticultural or agronomic utility, such as vegetable or other crops, can be transformed according to the present invention using techniques available in the art. A plant suitable for use in the present methods is Nicotiana tabacum, or tobacco. Any strain or variety of tobacco may be used. Additional plants (both monocots and dicots) which may be employed in practicing the present invention include, but are not limited to, potato (Solanum tuberosum), soybean (Glycine max), tomato (Lycopersicon esculentum), peanuts (Arachis hypogaea), cotton (Gossypium hirsutum), green beans (Phaseolus vulgaris), lima beans (Phaseolus limensis), peas (Lathyrus spp.)cassava (Manihot esculenta), coffee (Cofea spp.), pineapple (Ananas comosus), citrus trees (Citrus

10

15

20

25

30

spp.), banana (Musa spp.), corn (Zea mays), oilseed rape (Brassica napus), wheat, oats, barley, rye and rice. Thus, an illustrative category of plants which may be used to practice aspects of the present invention are the dicots, and a more particular category of plants which may be used to practice the present invention are members of the family Solanacae.

The methods of the present invention can further be practiced with turfgrass, including cool season turfgrasses and warm season turfgrasses. Examples of cool season turfgrasses are Biuegrasses (Poa L.), such as Kentucky Bluegrass (Poa pratensis L.), rough Bluegrass (Poa trivialis L.), Canada Bluegrass (Poa compressa L.), Annual Bluegrass (Poa annua L.), Upland Bluegrass (Poa glaucantha Gaudin), Wood Bluegrass (Poa nemoralis L.), and Bulbous Bluegrass (Poa bulbosa L.); the Bentgrasses and Redtop (Agrostis L.), such as Creeping Bentgrass (Agrostis palustris Huds.), Colonial Bentgrass (Agrostis tenius Sibth.), Velvet Bentgrass (Agrostis canina L.), South German Mixed Bentgrass (Agrostis L.), and Redtop (Agrostis alba L.); the Fescues (Festuca L.), such as Red Fescue (Festuca rubra L.), Chewings Fescue (Festuca rubra var. commutata Gaud.), Sheep Fescue (Festuca ovina L.), Hard Fescue (Festuca ovina var. duriuscula L. Koch), Hair Fescue (Festuca capillata Lam.), Tall Fescue (Festuca arundinacea Schreb.), Meadow Fescue (Festuca elatior L.); the Rye grasses (Lolium L.), such as Perennial Ryegrass (Lolium perenne L.), Italian Ryegrass (Lolium multiflorum Lam.); the Wheatgrasses (Agropyron Gaertn.), such as Fairway Wheatgrass (Agropyron cristatum L. Gaertn.), Western Wheatgrass (Agropyron smithii Rydb.). Examples of warm season turfgrasses are the Bermudagrasses (Cynodon L.C. Rich), the Zoysiagrasses (Zoysia Willd.), St. Augustinegrasses (Stenotaphrum secundatum (Walt.) Kuntze), Centipedegrass (Eremochioa ophiuroides (Munro.) Hack.), Carpetgrass (Axonopus Beauv.), Bahiagrass (Paspalum notatum Flugge.), Kikuyugrass (Pennisetum clandestinum Hochst. ex Chiov.), Buffalograss (Buchloe dactyloides (Nutt.) Engelm.), Blue Grama (Bouteloua gracilis (H.B.K.) Lag. ex Steud.), Sideoats Grama (Bouteloua curtipendula (Michx.) Torr.), and Dichondra

10

15

20

25

30

(Dichondra Forst.).

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a vector of the present invention. The term "organogenesis," as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers; the term "embryogenesis," as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

Plants of the present invention may take a variety of forms. The plants may be chimeras of transformed cells and non-transformed cells; the plants may be clonal transformants (e.g., all cells transformed to contain the transcription cassette); the plants may comprise grafts of transformed and untransformed tissues (e.g., a transformed root stock grafted to an untransformed scion in citrus species). The transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, first generation (or T1) transformed plants may be selfed to provide homozygous second generation (or T2) transformed plants, and the T2 plants further propagated through classical breeding techniques. A dominant selectable marker (such as nptII) can be associated with the transcription cassette to assist in breeding.

As used herein, a crop comprises a plurality of plants of the same genus or species, planted together in an agricultural field. By "agricultural field" is meant a common plot of soil or a greenhouse. Thus, the present invention provides a method of producing a crop of plants having altered metabolism of chemical compounds (such as a phenylurea herbicide), and thus having altered

10

15

20

25

30

resistance to the chemical compound, compared to a crop of non-transformed plants of the same genus or species, or variety.

Where a crop comprises a plurality of transgenic plants with increased resistance to phenylurea compounds according to the present invention, such compounds may be used as post-emergent herbicides to control undesirable plant species. Accordingly, a method of using phenylurea compounds as post-emergent herbicides according to the present invention comprises planting a plurality of transformed plant seed (or transformed plants) with enhanced resistance to a phenylurea herbicide, and applying that herbicide to the field after the germination and emergence of at least some of said transformed plant seed (or following the planting of transformed plants). Application of the phenylurea herbicide will selectively impact non-resistant plants.

9. Microbial decontamination:

Microbial cells useful for degrading phenylurea compounds, which cells contain and express a heterologous DNA molecule encoding a P-450 enzyme that enhances the metabolism of the phenylurea compound in the microbial cell (e.g., a peptide of SEQ ID NO:2), are a further aspect of the present invention. Suitable host microbial cells include soil microbes (i.e., those which grow in the soil) transformed to express a P-450 enzyme that enhances the metabolism of one or more phenylurea compounds by the host cell. Suitable microbes include bacteria (such as Agrobacterium, Bacillus, Streptomyces, Nocardia, etc.), fungi (including yeasts), and algae. Microbes can be selected, by methods known in the art of soil microbiology, to correspond to those which are typically found in the substrate to be treated. Liquids which are contaminated with phenylurea compounds may be contacted to transformed microorganisms by passing the contaminated liquid through a bioreactor which contains the microorganism. Numerous suitable bioreactor designs are known in the art. A microbial host particularly suitable for bioreactors is yeast.

Combination treatments utilizing aspects of the present invention involve

WO 99/19493 PCT/US98/20807

-20-

the application of a phenylurea compound in a location such as an agricultural field (e.g., as a herbicide), and subsequent application of a transformed microbe as described above in an amount effective to degrade residual applied herbicide. Application of the herbicide may be carried out in accordance with known techniques.

The examples which follow are set forth to illustrate the present invention, and are not to be construed as limiting thereof.

EXAMPLE 1

10

5

Materials and Methods

a. Substrates

Phenyl-U-[14C] fluometuron, phenyl-U-[14C] chlortoluron, phenyl-U-[14C] metolachlor, phenyl-U-[14C] prosulfuron, pyrimidinyl-2- diazinon, and phenyl-U-[14C] alachlor were provided by Novartis (Greensboro, North Carolina); phenyl-U-[14C] bentazon was donated by BASF (Research Triangle Park, North Carolina); phenyl-U-[14C] linuron, phenyl-U-[14C] diuron, and carbonyl-[14C] metribuzin were a gift from DuPont de Nemours (Wilmington, Delaware); carboxyl-[14C] imazaquin was provided by American Cyanamid (Princeton, New Jersey).

20

25

30

15

b. Isolation of P-450 cDNAs

Random amplification of partial cDNAs encoding P-450 enzymes was conducted essentially as described by Meijer et al., *Plant Mol. Biol.* 22:379-383 (1993), using a soybean (*Glycine max* cv Dare) leaf cDNA library as the template (Dewey et al., *Plant Cell* 6:1495-1507 (1994)). Briefly, degenerate inosine-containing primers were synthesized based on the highly conserved heme-binding region. The precise sequences of these primers are described in Meijer et al. (1993). An oligo-dT primer complementary to the poly(A) tail of the cDNA clones was used in conjunction with the degenerate primers in PCR amplification assays. Amplification products were cloned into the T-tailed pCRII plasmid

10

15

20

25

(Invitrogen, San Diego, CA) and DNA sequence analysis of the first 300-400 base pairs downstream of the conserved region was used to establish whether a given amplification product represented a true P-450 cDNA.

To recover full-length versions of the partial cDNAs, a primer (5'-TGTCTAACTCCTTTCC-3') (SEQ ID NO:19) complementary to the pYES2 vector (the vector into which the soybean cDNA library was cloned) and a downstream primer corresponding to a segment of the 3' untranslated region for each of the unique P-450 cDNAs were used in PCR reactions using the same soybean cDNA library as the template. PCR products were again cloned into the pCRII plasmid and the entire DNA sequence was determined for the largest cDNA amplified for each unique soybean P-450.

To isolate full-length versions of the respective P-450 ORFs without including any of the 5' untranslated region (which has been shown to potentially impede gene expression in yeast (Pompon, Eur. J. Biochem. 177:285-293 (1988)), an additional PCR reaction was performed with two gene-specific The forward primers contained a BamHI restriction site immediately followed by the ATG start codon, and the next 14-15 bases of the reading frame; the downstream primer was again specific for the 3' untranslated regions of the respective genes and included sequences specifying either EcoRI, KpnI, and SacI to facilitate subcloning of the P-450 cDNAs into the yeast expression vector, pYeDP60 (V-60; Urban et al., Biochimie 72:463-472 (1990)).

All PCR reactions, with the exception of the initial amplification of the partial P-450 cDNAs (see Meijer et al. (1993)), contained 0.2 ng/µl template, 2 μM of each primer, 200 μM of each dNTP, and 1.5 mM MgCl₂ in a final reaction volume of 50 µl. Amplification was initiated by the addition of 1.5 U EXPAND™ High Fidelity enzyme mix using conditions described by the manufacturer (Boeringer Mannheim). DNA sequence was determined by the chain termination method (Sanger et al., Proc. Natl. Acad. Sci. USA 74:5463-5467 (1977)) using fluorescent dyes (Applied Biosystems, Foster City, CA).

DNA and predicted amino acid sequences were analyzed using the BLAST 30

10

15

20

25

30

algorithm and the GAP program (University of Wisconsin, Madison, Genetics Computing Group software package).

c. P-450 cDNA Expression in Yeast

Yeast transformation was performed as described by Geitz et al., *Nucleic Acids Research* 20:1425 (1992). Media composition, culturing conditions, galactose induction, and microsomal preparations were conducted according to Pompon et al., *Methods Enzymol*. 272:51-64 (1995), using a culture volume of 250 ml. Microsomal protein was quantified spectrophotometrically using the method of Waddell, *J. Lab. Clin. Med*. 48:311-314 (1956), using bovine albumin as a standard. Dithionite-reduced, carbon monoxide difference spectra was obtained as previously outlined (Estabrook and Werringloer, *Methods Enzymol*. 52:212-220 (1978)) using a Shimadzu Recording Spectrophotometer UV-240 (Shimadzu, Kyoto, Japan). P-450 protein concentrations of yeast microsomes were calculated using a millimolar extinction coefficient of 91 (Omura and Sato, *J. Biol. Chem.*, 239:2370-2378 (1964)).

d. In vitro Herbicide Metabolism Assays

Yeast microsomes enriched for a discrete soybean P-450 isozyme were assayed for their capacity to metabolize the ten herbicides and one insecticide listed in Table 3. The reaction mixtures contained 10,000 DPM (100-200 ng) radiolabeled substrate, 0.75 mM NAPDH, 2.5 mg/ml microsomal protein. Total reaction volumes were adjusted to 150 µl with 50 mM phosphate buffer (pH 7.1). The mixtures were incubated under light for 45 minutes at 27°C, arrested with 50 µl acetone and centrifuged at 14 000xg for 2 minutes. Fifty microliters of the supernatants containing radiolabeled alachlor, metolachlor, metribuzin, prosulfuron, chlortoluron, diuron, fluometuron, linuron, or diazinon were spotted onto 250 micron Whatman K6F silica plates. Radiolabeled bentazon and imazaquin-containing samples were spotted onto 200 micron Whatman LKC18F silica gel reversed-phase plates. All plates were developed in a benzene/acetone

10

15

2:1 (v/v) solvent system with the exception of prosulfuron, developed in toluene/acetic acid, 75:20:5 (v/v/v), and bentazon and imazaquin, developed in methanol/75 mM sodium acetate 40:60 (v/v). The developed plates were scanned with a Bioscan System 400 imaging scanner (Bioscan, Washington, DC), and the production of metabolites was determined based on the chromatographic profiles. For microsomes containing the expressed CYP71A10 enzyme, control experiments were also conducted to measure the NADPH-dependency, and the inhibitory effects of CO. CO treatment of the sample was achieved by gentle bubbling of the gas through the reaction mixture for 2 minutes immediately before the assay was initiated by the addition of NADPH.

e. Enzyme Kinetics

Substrate conversion was quantified by a combination of TLC analysis and scintillation spectrometry. The location of the metabolic products on the TLC plates was identified using an imaging scanner, the bands were scraped and analyzed by scintillation spectrometry. The amount of metabolite produced was calculated based on specific activity and scintillation counts. Each assay was repeated at least twice. K_m and V_{max} values were estimated using nonlinear regression analysis.

20

f. Mass Spectral Analysis

The reaction components used in the *in vitro* fluometuron and linuron metabolism assays were scaled up 50-fold, and the reactions were allowed to proceed for 3 hours. The substrates and the metabolites were extracted 3 times with 20 ml ethyl acetate. The extracts were combined, evaporated to dryness, and the resulting pellet was resuspended in 1 ml acetone. The samples were purified twice using preparative TLC and imaging scanning as described above. Finally, the respective bands were scraped, the compounds were eluted with acetone and flash evaporated.

30

25

Fractions of interest were analyzed by liquid chromatography/mass

10

15

20

30

spectrometry (LC/MS). Mass spectral measurements were made with a Finnigan TSQ 7000 triple quadruple mass spectrometer (QQQ) equipped with an Atmospheric Pressure Ionization (API) interface fitted with a pneumatically assisted electrospray head (Finnigan MAT, Brennan, Germany). The spray nozzle was operated at 5 kV in the positive ion mode and 4 kV in the negative ion mode. For sample introduction, the TSQ 7000 was equipped with a HPLC solvent delivery system (Perkin-Elmer 410 LC pump), a UV detector (Perkin-Elmer), a stream splitter set at 6:1 with the majority of the effluent flowing to a radioisotope flow monitor (IN/US β-RAM) and the other stream attached to the API interface. Samples were chromatographed on a reverse phase HPLC column (Inertsil 5 ODS2, 150 x 2 mm i.d.). The column was eluted at 0.4 ml/min with 95:5 of 0.1% trifluoroacetic acid in water and 0.1% trifluoroacetic acid in methanol, respectively. Collision induced dissociation experiments (MS/MS) were conducted using argon gas with collision energy in the range of 17.5-30 eV at cell pressures of approximately 0.28 Pa. Signals were captured using a Finnigan 7000 data system.

g. NMR Analysis

Proton NMR measurements were made on a Bruker AMX-400 NMR spectrometer equipped with either a QNP or inverse probe set at 400.13 MHZ. Spectra were acquired at ambient temperature in acetonitrile- d_3 . Chemical shifts were expressed as parts per million, relative to the resonance of residual acetonitrile protons at 1.93 ppm (δ).

25 h. Tobacco Transformation

A plant expression vector capable of mediating the constitutive expression of CYP71A10 was produced. The GUS open reading frame of the binary expression vector pBI121 (Clontech, Polo Alto, CA) was excised and replaced with the full length CYP71A10 reading frame. This placed the soybean gene under the transcriptional control of the strong constitutive CaMV 35S promoter.

10

15

20

The resulting construct was used to transform Agrobacterium tumefaciens strain LBA 4404 (Holsters et al., Mol. Gen. Genetics, 163:181-187 (1988)). Excised leaf discs of Nicotiana tabacum cv SR1 were transformed using the Agrobacterium, and kanamycin-resistant plants were selected as described by Horsch et al. Science, 227:1229-1231 (1985). Primary transformants were potted in a standard soil mixture, transferred to a greenhouse and their seed harvested upon maturation.

i. In vivo Herbicide Metabolism Assays

Seeds from primary transgenic tobacco plants transformed with CYP71A10 and control plants transformed with the pBI121 vector were grown in Petri dishes containing MS salts and 100 µg/ml kanamycin. At five weeks post-seeding, kanamycin-resistant plantlets were transplanted into pots containing soil and grown an additional two weeks. Single leaves of approximately 10 cm² in size were excised and their petioles inserted into 100 µl of H₂O containing radiolabeled herbicide. The leaves were placed in a growth chamber maintaining a temperature of 27°C and incubated until the entire volume of the herbicide solution was drawn up by the transpirational stream of the leaves (about 3 hrs). The leaves were subsequently transferred into an Eppendorf tube containing distilled water and further incubated for a total of 14 hours.

[14C]-labeled herbicide was extracted from the leaves by grinding for 5 minutes in 250 μl methanol with a plastic pellet pestle driven by an electric drill. After centrifugation for 3 minutes at 14,000 g, 75 μl of the supernatant was spotted on a Whatman K6F silica plate and developed in a solvent system containing chloroform/ethanol/acetic acid 135:10:15 (v/v/v). The separated herbicide derivatives were visualized using an imaging scanner. Substrate conversion was quantified based on the amount of herbicide absorbed, and the ratios of the parent compound and the produced metabolites determined from the TLC profiles.

30

10

15

20

25

30

i. Herbicide Tolerance

T₁ generation seeds from CYP71A10-transformed tobacco and pBI121-transformed control plants were placed onto Petri dishes containing MS salts and linuron (using its commercial formulation LOROX 50 DF) at active ingredient concentrations ranging from 0.25 to 3.0 μM. Chlortoluron was added at 0, 1.0, 5.0 and 10.0 μM concentrations using a 99.5% pure analytical standard. The Petri dishes were incubated in a growth chamber maintaining a constant temperature of 27°C and a 16/8 hour light/dark cycle. The phytotoxic effects of the treatments were determined visually by comparison to control plants and plants grown in the absence of the herbicide. All treatments were repeated at least twice.

EXAMPLE 2

Isolation of P-450 cDNAs

To isolate cDNAs encoding P-450s from soybean, the PCR strategy described by Meijer et al. (1993) was adapted, using a soybean leaf cDNA library as the template. Degenerate, inosine-containing PCR primers were constructed corresponding to the first nine codons encoding the conserved sequence FLPFGxGxRxCxG (x = any amino acid) (SEQ ID NO:20), which represents an extension of the highly conserved FxxGxxxCxG motif (Bozak et al., *Proc. Natl. Acad. Sci. USA* 87:3904-3908 (1990)) (SEQ ID NO:21). Located near the C-terminal end of the protein, this motif defines the hemebinding region of the protein and may be regarded as a "signature" for P-450 proteins. A second nonspecific primer complementary to the poly(A) tail of the cDNA clones was used in conjunction with these degenerate primers in a PCR amplification assay. PCR amplification products were cloned into a plasmid vector and analyzed by DNA sequencing. Of 86 randomly selected individuals that were sequenced, 15 clones representing 10 unique cDNAs were identified that possessed the conserved cysteine and glycine residues of the signature

10

15

20

25

30

consensus (xCxG) (SEQ ID NO:22) immediately following the sequence defined by the degenerate PCR primers. Furthermore, homology searches of the major DNA and protein data bases revealed additional sequence identities to previously reported P-450 sequences for each of the ten unique soybean sequences (data not shown). Because this strategy only allows the recovery of sequence corresponding to the C-terminal portion of the proteins, additional PCR-based techniques were utilized to obtain cDNAs possessing the entire reading frames for each clone. Full length cDNAs were isolated for eight of the 10 individual clones and a near full length cDNA was isolated for an additional clone.

The eight full length and one near full length soybean P-450 cDNAs isolated are described in Table 1. The nomenclature for P-450 genes is based on amino acid sequence identity. Typically, sequences sharing >40% identity are placed in the same family, >55% identity defines members of the same subfamily, and sequences that display >97% identity are assumed to represent allelic variants, although exceptions to these designations have been noted (Nelson et al., *Pharmacogenetics*, 6:1-41 (1996)). According to this system of nomenclature, all of the nine soybean cDNAs were able to be placed within existing P-450 gene families; however, three of the sequences (CYP82C1, CYP83D1 and CYP93C1) defined new subfamilies. Although an increasing number of P-450 gene products have been assigned specific enzymatic functions (reviewed in Schuler, 1996), none of the soybean cDNAs listed in Table 1 could be placed into families for which an *in vivo* function had been determined for any of its members.

In addition to the conserved heme-binding domain described previously, all of the predicted soybean polypeptides possess slight variations of the conserved sequence PEEFxPERF (SEQ ID NO:23) located approximately 30 amino acids forward of the heme-binding motif (Hallahan et al., *Biochem. Soc. Trans.* 21:1068-1073 (1993)). Also characteristic of microsomal P-450s is the presence of an N-terminal noncleavable signal sequence that serves as the membrane anchor. Immediately following this signal-anchor segment in most

microsomal P-450s is a proline-rich region that is believed to form a hinge between the catalytic cytoplasmic domain and the hydrophobic membrane anchor (Halkier, *Phytochemistry* 43:1-21 (1996)). All of the present clones (except CYP97B2) encode proteins possessing predicted signal sequences; all individuals (except CYP97B2 and CYP82C1) contain readily identifiable proline-rich domains following the signal sequence (Table 1). It is the identification of both of these N-terminal motifs in the CYP83D1 encoded protein (but no Met codon) that indicates that this clone is nearly full length. Interestingly, instead of possessing a predicted signal sequence and proline-rich region, the N-terminus of the polypeptide encoded by clone CYP97B2 contains a motif characteristic of a chloroplast transit peptide (data not shown).

Table 1
Soybean P-450s Isolated Using Degenerate PCR Primers

Name	GenBank Accession #	Length (amino acids)	Closest Match	Identity* %	Membrane Anchor	Proline -rich Region
CYP71A10 (SEQ ID NO:1)	AF022157	513	CYP71A1	51.7	+	+
CYP71D10 (SEQ ID NO:3)	AF022459	510	CYP71D9	50.9	+	+
CYP77A3 (SEQ ID NO:5)	AF022464	513	CYP77A1	69.8	+	+
CYP78A3 (SEQ ID NO:7)	AF022463	523	CYP78A2	53.1	+	+
CYP82C1 (SEQ ID NO:9)	AF022461	532	CYP82A3	51.1	+	
CYP83D1** (SEQ ID NO:11)	AF022460	516	CYP71A1**	45.7	+	+
CYP93C1 (SEQ ID NO:13)	AF022462	521	CYP93B1	44.5	+	+
CYP97B2 (SEO ID NO:15)	AF022457	576	CYP97B1	80.8	_	_
CYP98A2 (SEQ ID NO:17)	AF022458	509	CYP98A1	69.7	+	+

¹⁵

20

WO 99/19493

5

^{*}Percent identity between the predicted amino acids sequences of the given soybean P-450 cDNA and the closest match identified from a BLAST search against the major gene and protein databases.

^{**} Although this sequence shows a best match to CYP71A1, it matches poorly to some sequences of the CYP71B subfamily. As a result, the tree cluster program places it into the CYP83 family.

10

15

20

25

30

EXAMPLE 3

Expression of Soybean P-450 cDNAs in Yeast

Because superfluous 5' untranslated sequences from foreign genes have been shown to be capable of impeding gene expression in yeast (Pompon, 1988), an additional PCR reaction was performed on each clone that enabled the cloning of full length P-450 open reading frames (ORFs) into the yeast expression vector pYeDP60 (V-60) without including any of the endogenous 5' nontranslated flanking sequence (see Methods). For the near full length clone CYP83D1, the 5' primer was also designed to generate an "artificial" Met start codon and a Val second codon at the 5' end of the ORF. Expression in yeast of genes cloned into the V-60 vector is mediated by the strong, galactose-inducible GAL10-CYC1 promoter (Pompon et al., 1995).

Previous studies have revealed that the heterologous expression of P-450 cDNAs in yeast can be greatly enhanced in strains that have been engineered to overexpress endogenous NADPH-dependent cytochrome P-450 reductase (Pompon et al., 1995). In strain W(R), this was accomplished by exchanging the relatively weak endogenous cytochrome P-450 reductase promoter with the same GAL10-CYC1 promoter used in vector V-60 (Truan et al., Gene 125:49-55 (1993)). To maximize the heterologous expression of the soybean P-450 cDNAs in yeast, each of the constructs cloned into the V-60 vector was transformed into strain W(R) and microsomes were isolated from cultures that had been induced by galactose.

Reduced-CO difference spectroscopy provides a method to measure the effectiveness of expression of heterologous P-450s in yeast. Microsomal preparations corresponding to five of the soybean constructs (CYP71A10, CYP71D10, CYP77A3, CYP83D1 and CYP98A2) showed characteristic P-450 CO difference spectra with Soret peaks at 450 nm; the profile corresponding to CYP71A10 is shown in Figure 1. No such peaks were observed for the remaining four clones. The specific P-450 content of the five positive

microsomal preparations varied significantly, ranging from 11 pmol P-450/mg protein for construct CYP83D1 to 252 pmol P-450/mg for clone CYP77A3 as shown in Table 2.

Table 2
P-450 Content of Microsomes Isolated from Yeast Overexpressing Various
Soybean CYPs

Clone	CYP content (pmol mg ⁻¹ protein)					
CYP71A10	44					
CYP71D10	15					
CYP77A3	252					
CYP83D1	11					
CYP98A2	13					

10

15

20

25

5

EXAMPLE 4

In vitro Herbicide Assays

To determine whether any of the present soybean P-450 proteins synthesized in yeast displayed significant herbicide metabolic activity, microsomal preparations possessing each of the five soybean P-450s that were effectively expressed in yeast (as judged by their reduced CO difference spectra, see above) were incubated individually with NADPH and radioisotopes of the compounds listed in Table 3. These substrates represent six different classes of herbicides and one organophosphate insecticide (diazinon). Upon termination of the reaction, each sample was analyzed by thin layer chromatography (TLC) to reveal potential metabolic breakdown products.

The P-450 proteins expressed from clones CYP71D10, CYP77A3, CYP83D1, and CYP98A2 displayed no apparent *in vitro* metabolic activity against any of the 11 compounds tested (data not shown). In contrast, the P-450 enzyme produced from construct CYP71A10 demonstrated considerable activity

uš.

5

10

15

20

25

against the phenylurea class of herbicides, but no activity against the remaining compounds. As shown in Figure 2, fluometuron and diuron were converted to a single metabolite; linuron and chlortoluron were transformed into two (a major and a minor) metabolites. Figure 3 shows the chemical structures of the four phenylurea herbicides tested in this study, and the derivatives that have previously been characterized as the first metabolites produced during the detoxification of the respective herbicides in plants known to metabolize these compounds (Voss and Geissbühler, *Proc. Brit. Weed Contr. Conf.* 8:266-268 (1966); Suzuki and Casida, *J. Agric. Food Chem.* 29:1027 (1981); Ryan et al., *Pestic. Biochem. Physiol.* 16:213-221 (1981)).

To further confirm that the herbicide metabolism measured from microsomes of yeast expressing CYP71A10 was attributable to a P-450 activity, additional assays utilizing linuron as the substrate were conducted. As shown in Figure 4, linuron metabolizing activity is reduced approximately 37% in the presence of CO, and no metabolites are observed when NADPH is omitted from the reaction. Activity is also completely abolished upon addition of tetcyclasis, a potent P-450 inhibitor (data not shown). Furthermore, no activity is detected when microsomal preparations are used from yeast cells expressing only the V-60 control plasmid. These results verify that the observed herbicide metabolizing activity is derived from the soybean CYP71A10 cDNA.

The kinetic properties and catalytic activities of the soybean CYP71A10 protein enzyme differed significantly among the four phenylurea-type herbicide substrates. As shown in Table 4, turnover rates for fluometuron and linuron were considerably greater than those observed for chlortoluron and diuron. The observed reduced activity for the later two substrates is apparently not the result of decreased binding affinities since the apparent K_ms for chlortoluron and diuron are lower than those measured for fluometuron and linuron.

Table 3

30 Compounds Used in Metabolism Assays

Common Name	Chemical Family
Alachlor	Acetanilide
Metolachlor	Acetanilide
Bentazon	Benzothiadiazole
Imazaquin	Imidazolinone
Chlortoluron	Phenylurea
Diuron	Phenylurea
Fluometuron	Phenylurea
Linuron	Phenylurea
Prosulfuron	Sulfonylurea
Metribuzin	as-Triazine
Diazinon	Organophosphate

WO 99/19493 PCT/US98/20807

-33-

Table 4
In Vitro Kinetic Parameters of the CYP71A10 Enzyme for Four Phenylurea Substrates

	K	V _{max}	Turnover		
Substrate	(µM)	(pmol min ⁻¹ mg ⁻¹ protein)	(min ⁻¹)		
Fluometuron	14.9 (1.0)*	303.6 (10.8)	6.8 (0.24)		
Linuron	9.8 (2.1)	125.6 (12.0)			
Chlortoluron	1.0 (0.2)	29.4 (2.2)	0.7 (0.05) 0.4 (0.04)		
Diuron	1.5 (0.3)	16.8 (1.6)			

- 5 * Values in parentheses represent standard error.
 - Assays were repeated three times for linuron and twice for all other substrates.
 - Concentration ranges (μM) used were 3.2-27.7 for fluometuron, 3.8-28.3 for linuron, 0.7-4.0 for chlortoluron, and 0.7-3.7 for diuron.

10

15

20

25

EXAMPLE 5

Analysis of Metabolites

As shown in Figure 2, CYP71A10-mediated degradation of phenylurea herbicides resulted in the accumulation of either one or two metabolites, depending on the particular substrate used. To determine the structure of the metabolites, the single metabolite observed in the fluometuron assay and both the major and minor metabolites generated in the linuron assay were analyzed by liquid chromatography/mass spectroscopy (LC/MS) analysis (results not shown). Analysis of the fluometuron metabolite by LC/MS in positive ion mode resulted pseudomolecular ions at m/z 219 [(M+H) $^+$, C₉H₉F₃N₂O] and m/z 241 $(M+Na)^+$ that corresponds to a sodium adduct. Daughter ion spectra of m/z 219 produced a prominent m/z 162 fragment ion due to formation of the protonated trifluoromethylaniline $(C_7H_6F_3N+H)^+$. Analysis of the fluometuron metabolite by proton NMR showed a singlet at $\delta 2.71$ which integrated for 3 protons (data not shown). The NMR spectra aromatic resonances were similar to aromatic Spectra of the fluometuron resonances observed in the parent molecule. metabolite were consistent for loss of a methyl group from the parent compound.

10

15

20

25

The major linuron metabolite analyzed by LC/MS in the negative ion mode showed a pseudomolecular ion at m/z 233 (M-H) and m/z 235 [(M+2)-H] consistent for a molecule containing two chlorine atoms. Daughter ion spectrum at m/z 233 showed a prominent fragment ion at m/z 160 (C₆H₄Cl₂N-H). The major linuron metabolite was 15 mass units less than parent compound which is consistent with loss of a methyl group. The position of methyl loss could not be determined based on mass spectral data alone.

The minor linuron metabolite analyzed by LC/MS gave a pseudomolecular ion at m/z 217 (M-H) and m/z 219 [(M+2)-H] which is consistent for a molecule containing two chlorine atoms. The daughter ion spectrum at m/z 217 showed a prominent fragment ion at m/z 160 which corresponds to formation of the dichloroaniline. The mass spectral data is consistent for the minor linuron metabolite representing N-demethoxy linuron.

These results suggest that the CYP71A10 enzyme expressed in yeast produces the same fluometuron and linuron metabolites as depicted in Figure 3, which shows the first metabolites produced during the detoxification of the respective herbicides in plants that are known to degrade these compounds. The metabolites of chlortoluron and diuron have not been analyzed directly, but the R_f values of the peaks observed during TLC separation are consistent with these species also representing the compounds shown in Figure 3 (ring-hydroxymethyl chlortoluron, N-demethyl chlortoluron and N-demethyl diuron). These results indicate that the CYP71A10 enzyme functions primarily as an N-demethylase with respect to fluometuron, linuron and diuron, with some N-demethoxylase activity also observed with linuron. Using chlortoluron as a substrate, the enzyme apparently functions primarily as a methyl-ring hydroxylase and to a lesser extent as an N-demethylase.

EXAMPLE 6

Herbicide Metabolism in Transgenic Tobacco

To determine whether overexpression of the soybean CYP71A10 cDNA

10

15

20

25

30

in a higher plant system enhances metabolism of phenylurea herbicides, the GUS gene in the binary vector pBI121 was excised and replaced with the CYP71A10 reading frame. This construct placed the CYP71A10 cDNA under the transcriptional control of the constitutive 35S promoter of Cauliflower Mosaic Virus; kanamycin selection was facilitated via the nptII selectable marker. Agrobacterium mediated transformation of Nicotiana tabacum cv SR1 leaf discs resulted in the recovery of several dozen independent kanamycin-resistant transformants. The plants were subsequently grown to maturity in a greenhouse and allowed to set seed.

For the herbicide metabolism assays, seeds from one randomly selected transgenic line, designated 25/2, were germinated on kanamycin-containing media to eliminate potential nontransgenic segregants. Of 17 germinated seedlings grown, only one individual was inhibited by kanamycin (data not shown). This result suggests that line 25/2 possesses more than one independently segregating transgene. Individual leaves from the 25/2 progeny were excised and incubated with radiolabeled phenylurea herbicides. As shown in Table 5, leaves of the kanamycin-resistant individuals of line 25/2 metabolized all of the four herbicides tested to a much greater extent than the pBI121-transformed control plants.

The relative migrations of the metabolic products revealed by TLC suggest that the products observed in the *in vivo* excised leaf assay are primarily the same as were generated from the *in vitro* assays using yeast microsomes for fluometuron, linuron and diuron (data not shown). For chlortoluron, additional metabolites were also observed. These likely represent combinations of ringmethyl hydroxylated and mono- and di-demethylated species as had been observed by Shiota et al. *Pestic. Biochem. Physiol.* 54:190-198 (1996), in their analysis of chlortoluron-resistant transgenic tobacco that overexpressed the rat CYP1A1 gene. Differences in the ratios of the observed chlortoluron metabolites were also observed between the CYP71A10-transformed and the control plants. Sixty three percent of the metabolites produced in the control leaves was N-

demethyl chlortoluron; in contrast, ring-methyl hydroxy chlortoluron was the most abundant metabolite generated in the CYP71A10-transformed leaves (47%) and only 8% of the metabolites represented N-demethyl chlortoluron.

Table 5

Phenylurea Metabolism after 14 Hours by Excised Leaves of Transgenic

Tobacco Plant 25/2 Progeny

Herbicide ^a	CYP71A10-transformed	Control ⁵					
	% of herbicide metabolized						
Fluometuron	91 (4.5)°	15 (0.6)					
Linuron	87 (2.0)	12 (2.6)					
Chlortoluron	85 (8.1) ⁴	39 (7.5) ^d					
Diuron	49 (7.0)	20 (2.0)					
Dinion							

- (a) Equal amounts of herbicide (1.2 nmol) were added for each experiment.
- (b) Plants transformed with the pBI121 construct were used as controls.
- (c) Values in parentheses represent standard error. A single leaf was assayed from four independent 25/2 plants and three independent control plants.
- (d) The major chlortoluron metabolite in the control plants represented N-demethyl chlortoluron (63%). The metabolites recovered from the CYP71A10-transformed leaves were ring-methyl hydroxy chlortoluron (47%), N-demethyl chlortoluron (8%) and other derivatives (45%).

20

25

15

10 :

EXAMPLE 7

Herbicide Tolerance

To establish whether enhanced herbicide metabolism leads to an increase in tolerance at the whole plant level, seeds from transgenic plant 25/2 were germinated on an agarose-base medium containing MS salts and varying

10

15

20

25

30

concentrations of linuron. Growth of wild-type SR1 plants and transgenic control plants expressing the GUS gene (from vector pBI121) was severely inhibited when exposed to 0.25 µM linuron and completely arrested at concentrations of 0.5 µM and higher (data not shown). As shown in Figure 5, progeny of plant 25/2 grown on media containing no herbicide (Figure 5A) appeared indistinguishable from the same seed grown in the presence of 0.5 μM linuron (Figure 5C), where only one of 23 germinated seedlings appeared to be inhibited by the herbicide. This ratio appears to be consistent with that observed when seeds from the same parent were grown on selective media containing kanamycin; only one of 17 seedlings failed to grow in the presence of kanamycin. Figure 5B shows control tobacco plants (transformed with vector pBI121), grown on media containing 0.5µM linuron. 25/2 plants tolerant to linuron levels as high as $2.5 \mu M$ linuron were observed, although an increasing percentage of the plants showed growth inhibition as the herbicide concentration was increased (Figure 5D). Segregation of the transgene(s) may be leading to variability in expression levels among the progeny of 25/2.

To examine whether the acquisition of herbicide tolerance is unique to line 25/2, seeds from 20 other independent CYP71A10-expressing transgenic plants were similarly germinated and grown on media containing 0.5 μ M linuron. Of these, 19 lines gave rise to progeny that were linuron tolerant. The percentage of tolerant individuals for each line varied from approximately 20% to 100% (data not shown). This variation likely represents differences in the copy number, expression levels and segregation of the transgene among the independent lines.

Chlortoluron-tolerance of line 25/2 was also evident. At 1.0 µM herbicide concentration chlortoluron completely arrested the growth of the control plants (Figure 5E). Although growth of the 25/2 plants was modestly inhibited at this herbicide concentration, with the exception of two presumably nontransgenic segregants, the CYP71A10-transformed plants appeared healthy (Figure 5F). In contrast to linuron and chlortoluron, little tolerance of line 25/2

to fluometuron or diuron was observed. Herbicide concentrations that were injurious to the control plants also inhibited the growth of line 25/2 individuals. Enhanced fluometuron or diuron tolerance was only observed at the very lowest herbicide concentrations necessary to impose growth inhibition in the control plants (data not shown).

10

THAT WHICH IS CLAIMED IS:

- 1. An isolated DNA molecule comprising a sequence selected from the group consisting of:
 - a) SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, and SEQ ID NO:17;
 - b) DNA sequences which encode an enzyme having a sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, and SEQ ID NO:18;
 - c) DNA sequences which have at least about 90% sequence identity to the DNA of (a) or (b) above and which encode a cytochrome P450 enzyme; and
 - d) DNA sequences which differ from the DNA of (a) or (c) above due to the degeneracy of the genetic code.
 - 2. A peptide encoded by a DNA sequence of claim 1.
- 3. A cytochrome p450 enzyme having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, and SEQ ID NO:18.
- 4. An isolated DNA molecule comprising a sequence selected from the group consisting of:
 - a) SEQ ID NO:1;
 - b) DNA sequences which encode an enzyme having SEQ ID NO:2,;
 - c) DNA sequences which have at least about 90% sequence identity to the DNA of (a) or (b) above and which encode a cytochrome P450 enzyme; and

- d) DNA sequences which differ from the DNA of (a) or (c) above due to the degeneracy of the genetic code.
 - 5. A peptide encoded by a DNA sequence of claim 4.
 - 6. A cytochrome p450 peptide having SEQ ID NO:2.
 - 7. A DNA construct comprising an expression cassette, which construct comprising in the 5' to 3' direction, a promoter operable in a plant cell and a DNA segment according to claim 1 positioned downstream from said promoter and operatively associated therewith.
 - 8. A DNA construct according to claim 7, wherein said promoter is constitutively active in plant cells.
 - 9. A DNA construct according to claim 7, wherein said promoter is the 35S promoter from Cauliflower Mosaic virus.
 - 10. A DNA construct according to claim 7, said construct further comprising a plasmid.
 - 11. A DNA construct according to claim 7 carried by a plant transformation vector.
 - 12. A DNA construct according to claim 7 carried by an Agrobacterium tumefaciens plant transformation vector.
 - 13. A plant cell containing a DNA construct according to claim 7.
 - 14. A transgenic plant comprising plant cells according to claim 13.

- 15. A transgenic plant according to claim 14, wherein said plant is a monocot.
- 16. A transgenic plant according to claim 14, wherein said plant is a dicot.
- 17. A DNA construct comprising an expression cassette, which construct comprising in the 5' to 3' direction, a promoter operable in a plant cell, and a DNA segment encoding a peptide of SEQ ID NO:2 positioned downstream from said promoter and operatively associated therewith.
- 18. A DNA construct according to claim 17, wherein said promoter is constitutively active in plant cells.
- 19. A DNA construct according to claim 17, wherein said promoter is the 35S promoter from Cauliflower Mosaic virus.
- 20. A DNA construct according to claim 17, said construct further comprising a plasmid.
- 21. A DNA construct according to claim 17 carried by a plant transformation vector.
- 22. A DNA construct according to claim 17 carried by an Agrobacterium tumefaciens plant transformation vector.
 - 23. A plant cell containing a DNA construct according to claim 17.
 - 24. A transgenic plant comprising plant cells according to claim 23.

- 25. A transgenic plant according to claim 24, wherein said plant is a monocot.
- 26. A transgenic plant according to claim 24, wherein said plant is a dicot.
- 27. A method of making a transgenic plant cell having an increased ability to metabolize phenylurea compounds compared to an untransformed plant cell, said method comprising:
 - a) providing a plant cell;
 - b) transforming said plant cell with an exogenous DNA construct comprising, in the 5' to 3' direction, a promoter operable in a plant cell and a DNA sequence encoding a peptide of SEQ ID NO:2, said DNA sequence operably linked to said promoter.
- 28. A method according to claim 27, wherein said plant cell is from a member of the Solanacae family.
- 29. A method according to claim 27, wherein said promoter is the 35S promoter from Cauliflower Mosaic virus.
- 30. A method according to claim 27, wherein said transforming step is carried out by bombarding said plant cell with microparticles carrying said DNA construct.
- 31. A method according to claim 27 wherein said transforming step is carried out by infecting said plant cell with an *Agrobacterium tumefaciens* containing a Ti plasmid carrying said DNA construct.
- 32. A method according to claim 27, further comprising regenerating a plant from said transformed plant cell.

- 33. A transformed plant produced by the method of claim 32.
- 34. Seed or progeny of a plant according to claim 33, which seed or progeny has inherited said DNA sequence encoding a peptide of SEQ ID NO:2.
- 35. A transformed plant produced by the method of claim 32, which plant has increased resistance to phenylurea herbicides compared to wild-type plants of the same species.
- 36. A transgenic plant having an increased ability to metabolize phenylurea compounds compared to an untransformed plant cell, said transgenic plant comprising transgenic plant cells containing an exogenous DNA construct comprising, in the 5' to 3' direction, a promoter operable in said plant cell, said promoter operably linked to a DNA sequence encoding a peptide of SEQ ID NO:2.
- 37. A transgenic plant according to claim 36, wherein said promoter is the 35S promoter from Cauliflower Mosaic virus.
- 38. A transgenic plant according to claim 36, wherein said plant is a dicot.
- 39. A transgenic plant according to claim 36, wherein said plant is a monocot.
- 40. A transgenic plant according to claim 36, wherein said plant is a member of the family Solanacae.
- 41. A transgenic plant according to claim 36, which plant is selected from the group consisting of tobacco, potato, tomato, corn, rice, cotton, soybean,

rape, wheat, oats, barley, rye and rice.

- 42. Progeny or seed of a plant according to claim 36, wherein said seed or progeny has inherited said DNA sequence encoding a peptide of SEQ ID NO:2.
- 43. A transformed plant according to claim 36, which plant has increased resistance to phenylurea herbicides compared to wild-type plants of the same species.
- 44. A crop comprising a plurality of plants according to claim 36 planted in an agricultural field.
- 45. A method of using a phenylurea herbicide as a post-emergence herbicide, comprising:
 - a) planting a crop according to claim 44;
 - b) applying to said crop a phenylurea herbicide.
- 46. A method according to claim 45, wherein said crop is selected from the group consisting of turfgrass, tobacco, potato, tomato, corn, rice, cotton, soybean, rape, wheat, oats, barley, rye and rice.
- 47. A method according to claim 45, wherein said herbicide is selected from the group consisting of fluometuron, linuron, chlortoluron and diuron.

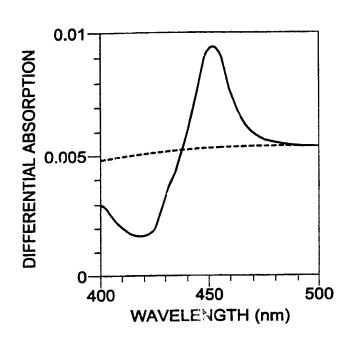
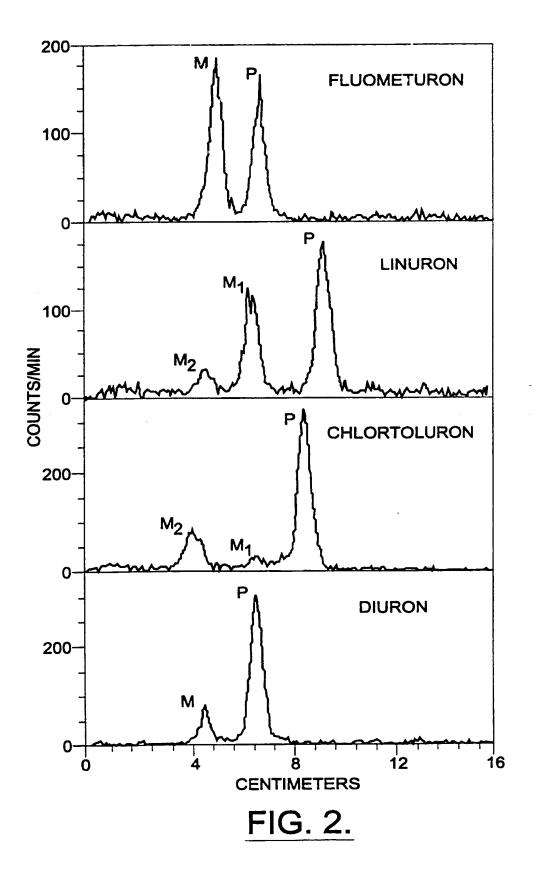


FIG. 1.

SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

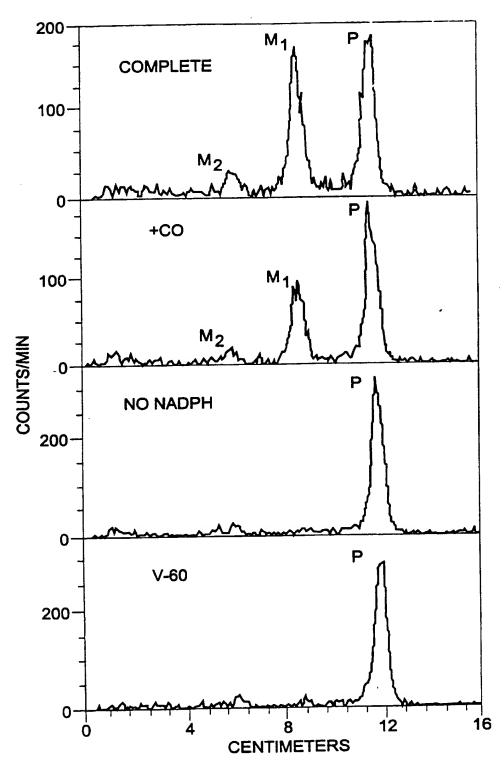


FIG. 4.

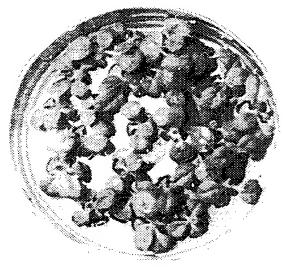
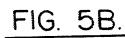
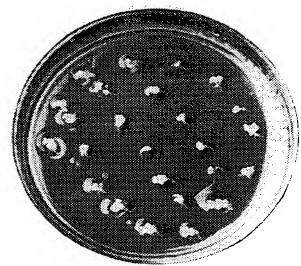


FIG. 5A.





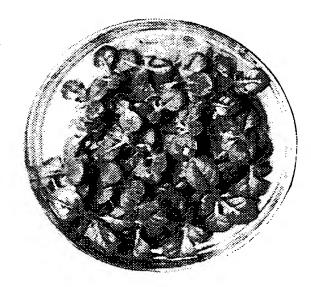
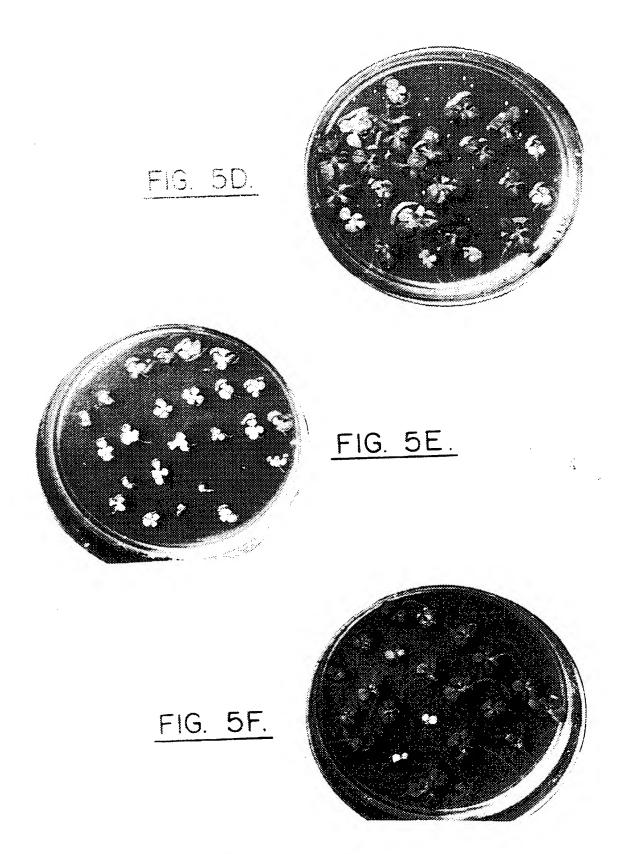


FIG. 5C.

SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

-1-SEQUENCE LISTING

- (1) CENERAL INFORMATION:
 - (i) APPLICANT: Siminszky, Balazs
 Dewey, Ralph E.
 Corbin, Frederick T.
 - (ii) TITLE OF INVENTION: Novel Cytochrome P-450 Constructs and Methods of Producing Herbicide-Resistant Transgenic Plants
 - (iii) NUMBER OF SEQUENCES: 23
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Virginia C. Bennett
 - (B) STREET: PO Box 37428
 - (C) CITY: Raleigh
 - (D) STATE: North Carolina
 - (E) COUNTRY: USA
 - (F) ZIP: 27627
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
 - (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Bennett, Virginia C.
 - (B) REGISTRATION NUMBER: 37,092(C) REFERENCE/DOCKET NUMBER: 5051-409
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 919-854-1400
 - (B) TELEFAX: 919-854-1401
 - (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1838 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 4..1542

672

720

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1: AAA ATG GCT CTA CTA TCA TCA GTC CTA AAG CAA TTG CCG CAT GAG CTA 48 Met Ala Leu Leu Ser Ser Val Leu Lys Gln Leu Pro His Glu Leu AGT TCA ACC CAT TAC CTA ACA GTT TTC TTC TGC ATC TTC CTT ATA CTT 96 Ser Ser Thr His Tyr Leu Thr Val Phe Phe Cys Ile Phe Leu Ile Leu 20 CTT CAG CTA ATA AGA AGA AAC AAA TAC AAT CTG CCA CCA TCC CCA CCA 144 Leu Gln Leu Ile Arg Arg Asn Lys Tyr Asn Leu Pro Pro Ser Pro Pro 40 AAG ATA CCC ATA ATC GGC AAT CTT CAC CAG CTA GGC ACA CTG CCA CAC 192 Lys Ile Pro Ile Ile Gly Asn Leu His Gln Leu Gly Thr Leu Pro His 55 CGC TCC TTT CAT GCA CTC TCA CAC AAA TAT GGC CCT CTC ATG ATG TTG . 240 Arg Ser Phe His Ala Leu Ser His Lys Tyr Gly Pro Leu Met Met Leu CAA TTG GGT CAA ATT CCA ACC CTA GTG GTC TCA TCA GCT GAC GTG GCC 288 Gln Leu Gly Gln Ile Pro Thr Leu Val Val Ser Ser Ala Asp Val Ala 90 85 80 AGA GAA ATA ATC AAA ACG CAT GAT GTT TTC TCC AAC CGC CGA CAA 336 Arg Glu Ile Ile Lys Thr His Asp Val Val Phe Ser Asn Arg Arg Gln 105 100 CCT ACA GCT GCT AAA ATC TTT GGT TAT GGA TGC AAA GAT GTG GCT TTC 384 Pro Thr Ala Ala Lys Ile Phe Gly Tyr Gly Cys Lys Asp Val Ala Phe GTG TAC TAC CGC GAA GAG TGG AGA CAA AAG ATA AAG ACA TGT AAG GTT 432 Val Tyr Tyr Arg Glu Glu Trp Arg Gln Lys Ile Lys Thr Cys Lys Val 135 130 GAG CTT ATG AGT CTG AAG AAG GTG CGG TTG TTT CAT TCC ATT AGA CAA 480 Glu Leu Met Ser Leu Lys Lys Val Arg Leu Phe His Ser Ile Arg Gln 150 145 GAA GTT GTT ACA GAG TTG GTT GAA GCT ATA GGT GAA GCG TGT GGT AGT 528 Glu Val Val Thr Glu Leu Val Glu Ala Ile Gly Glu Ala Cys Gly Ser 170 160 GAA AGA CCA TGT GTG AAT CTG ACT GAG ATG CTG ATG GCA GCA TCG AAC 576 Glu Arg Pro Cys Val Asn Leu Thr Glu Met Leu Met Ala Ala Ser Asn

GAC ATT GTG TCT AGA TGT GTT CTT GGA CGG AAG TGT GAT GAT GCA TGT

Asp Ile Val Ser Arg Cys Val Leu Gly Arg Lys Cys Asp Asp Ala Cys 200

GGT GGT AGT GGC AGT AGC AGC TTT GCA GCG TTG GGA AGA AAG ATT ATG

Gly Gly Ser Gly Ser Ser Ser Phe Ala Ala Leu Gly Arg Lys Ile Met 215

AGA CTA TTA TCG GCT TTC AGC GTG GGT GAT TTC TTC CCT TCG TTG GGT

180

WO 99/19493 PCT/US98/20807

Arg	Leu 225 _.	Leu	Ser	Ala	Phe	Ser 230	Val	Gly	-3 - Asp	Phe	Phe 235	Pro	Ser	Leu	Gly		
TGG Trp 240	GTT Val	GAC Asp	ТАТ Туг	CTG Leu	ACT Thr 245	GGC Gly	TTA Leu	ATT Ile	CCA Pro	GAG Glu 250	ATG Met	AAA Lys	ACC Thr	ACG Thr	TTT Phe 255		768
CTC Leu	GCA Ala	GTA Val	GAT Asp	GCT Ala 260	TTC Phe	CTT Leu	GAT Asp	GAG Glu	GTA Val 265	ATT Ile	GCA Ala	GAA Glu	CAC His	GAG Glu 270	AGC Ser		816
AGT Ser	AAC Asn	AAG Lys	AAG Lys 275	AAT Asn	GAT Asp	GAC Asp	TTC Phe	TTG Leu 280	GGG	ATA Ile	CTT Leu	CTT Leu	CAA Gln 285	CTT Leu	CAA Ģln		864
GAA Glu	TGT Cys	GGG Gly 290	AGG Arg	CTT Leu	GAC Asp	TTT Phe	CAG Gln 295	CTC Leu	GAC Asp	CGA Arg	GAT Asp	AAC Asn 300	CTC Leu	AAA Lys	GCA Ala		912
ATC Ile	CTA Leu 305	GTG Val	GAC Asp	ATG Met	ATA Ile	ATA Ile 310	GGT Gly	GGG Gly	AGT Ser	GAC Asp	ACT Thr 315	ACT Thr	TCA Ser	ACA Thr	ACT Thr		960
CTA Leu 320	GAA Glu	TGG Trp	ACT Thr	TTT Phe	GCG Ala 325	GAG Glu	TTC Phe	CTT Leu	AGA Arg	AAT Asn 330	CCA Pro	AAT Asn	ACC Thr	ATG Met	AAG Lys 335	3	1008
AAA Lys	GCT Ala	CAA Gln	GAA Glu	GAG Glu 340	GTA Val	AGA Arg	AGA Arg	GTG Val	GTG Val 345	GGA Gly	ATC Ile	AAT Asn	TCC Ser	AAA Lys 350	GCA ⁻ Ala	1	
GTA Val	CTG Leu	GAT Asp	GAA Glu 355	AAT Asn	TGT Cys	GTG Val	AAT Asn	CAA Gln 360	ATG Met	AAC Asn	TAC	TTG Leu	AAA Lys 365	TGT Cys	GTA Val	3	1104
GTC Val	AAA Lys	GAA Glu 370	ACT Thr	TTG Leu	AGA Arg	TTA Leu	CAT His 375	CCA Pro	CCC Pro	CTT Leu	CCT Pro	CTT Leu 380	TTG Leu	ATT Ile	GCT Ala	3	1152
CGA Arg	GAG Glu 385	ACA Thr	TCA Ser	TCA Ser	AGT Ser	GTA Val 390	AAA Lys	CTA Leu	AGA Arg	GGG Gly	TAC Tyr 395	GAT Asp	ATT Ile	CCC Pro	GCA Ala	3	1200
AAA Lys 400	ACA Thr	ATG Met	GTA Val	TTT Phe	ATC Ile 405	AAT Asn	GCA Ala	TGG Trp	GCG Ala	ATC Ile 410	CAG Gln	AGG Arg	GAT Asp	CCT Pro	GAA Glu 415	:	1248
TTA Leu	TGG Trp	GAT Asp	GAT Asp	CCT Pro 420	GAA Glu	GAA Glu	TTT Phe	ATT Ile	CCC Pro 425	GAA Glu	AGA Arg	TTT Phe	GAA Glu	ACT Thr 430	AGC Ser	:	1296
CAA Gln	GTT Val	GAT Asp	CTT Leu 435	AAT Asn	GGA Gly	CAA Gln	GAT Asp	TTT Phe 440	CAA Gln	TTA Leu	ATT	CCG Pro	TTC Phe 445	GGT Gly	ATT Ile	:	1344 .
GGG Gly	AGA Arg	AGG Arg 450	Gly	TGC Cys	CCT Pro	GCA Ala	ATG Met 455	Ser	TTT Phe	GGA Gly	CTT Leu	GCT Ala 460	TCA Ser	ACT Thr	GAG Glu	:	1392

	•																
TAT Tyr	GTT Val 465	Leu	GCT Ala	AAT Asn	CTT Leu	TTG Leu 470	TAT Tyr	TGG Trp	-4 TTC Phe	AAT	TGG Trp 475	AAT Asn	ATG Met	TCC	GAG Glu	144	0
TCT Ser 480	GGA Gly	CGT Arg	ATA Ile	TTG Leu	ATG Met 485	CAC His	AAC Asn	ATT Ile	GAC Asp	ATG Met 490	AGT Ser	GAG Glu	ACA Thr	AAT Asn	GGA Gly 495	148	8
CTC Leu	ACT Thr	GTC Val	AGT Ser	AAG Lys 500	AAA Lys	GTA Val	CCA Pro	CIT Leu	CAT His 505	CTT Leu	GAA Glu	CCA Pro	GAA Glu	CCA Pro 510	TAT Tyr	153	6
	ACA Thr	TGA	rcat'	TTC F	ACAT:	ratg(CA TO	TTT(GGCA.	A CAC	CCTAT	raaa	GAG'	rata:	ga Ť	159	12
CTG	gaag'	TAC :	TCA	ATTT	AG T	LAT G(GATG!	r aa	AAGC'	TATA	CAA	raag:	AAG 1	rgct)	AACAAG	165	;2
CTA	GGAT.	ATG A	AGCA'	TTTA'	rg G	AGTA	ACGA	G TG	AGGT'	TCCA	AAG	AGTC'	TAA '	rtac'	TCGTCT	171	. 2
CTT	GAAC.	ATT (GTTA'	TATT'	IG T	rttc'	TTGC	A GT	TTGT'	TAAT	CTT	TTGA	ATA (GTTG'	TTTCAC	177	12
ATT	TATT'	TTT (TATE	GGTT'	TG T	rggt:	ATGT	r GT	GGAA	GGCG	TTA	GTAA	AAA '	TTTG'	TGGTGT	183	32
GTT	CTT															183	38

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 513 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Leu Leu Ser Ser Val Leu Lys Gln Leu Pro His Glu Leu Ser

Ser Thr His Tyr Leu Thr Val Phe Phe Cys Ile Phe Leu Ile Leu Leu 20 25 30

Gln Leu Ile Arg Arg Asn Lys Tyr Asn Leu Pro Pro Ser Pro Pro Lys
35 40 45

Ile Pro Ile Ile Gly Asn Leu His Gln Leu Gly Thr Leu Pro His Arg
50 55 60

Ser Phe His Ala Leu Ser His Lys Tyr Gly Pro Leu Met Met Leu Gln 65 70 75 80

Leu Gly Gln Ile Pro Thr Leu Val Val Ser Ser Ala Asp Val Ala Arg 85 90 95

Glu Ile Ile Lys Thr His Asp Val Val Phe Ser Asn Arg Arg Gln Pro 100 105 110

Thr Ala Ala Lys Ile Phe Gly Tyr Gly Cys Lys Asp Val Ala Phe Val

-5-125 120 115 Tyr Tyr Arg Glu Glu Trp Arg Gln Lys Ile Lys Thr Cys Lys Val Glu 140 135 130 Leu Met Ser Leu Lys Lys Val Arg Leu Phe His Ser Ile Arg Cln Glu Val Val Thr Glu Leu Val Glu Ala Ile Gly Glu Ala Cys Gly Ser Glu 170 165 Arg Pro Cys Val Asn Leu Thr Glu Met Leu Met Ala Ala Ser Asn Asp 185 180 Ile Val Ser Arg Cys Val Leu Gly Arg Lys Cys Asp Asp Ala Cys Gly Gly Ser Gly Ser Ser Ser Phe Ala Ala Leu Gly Arg Lys Ile Met Arg 215 Leu Leu Ser Ala Phe Ser Val Gly Asp Phe Phe Pro Ser Leu Gly Trp Val Asp Tyr Leu Thr Gly Leu Ile Pro Glu Met Lys Thr Thr Phe Leu 250 Ala Val Asp Ala Phe Leu Asp Glu Val Ile Ala Glu His Glu Ser Ser Asn Lys Lys Asn Asp Asp Phe Leu Gly Ile Leu Leu Gln Leu Gln Glu 280 Cys Gly Arg Leu Asp Phe Gln Leu Asp Arg Asp Asn Leu Lys Ala Ile 295 Leu Val Asp Met Ile Ile Gly Gly Ser Asp Thr Thr Ser Thr Thr Leu 315 310 Glu Trp Thr Phe Ala Glu Phe Leu Arg Asn Pro Asn Thr Met Lys Lys Ala Gln Glu Glu Val Arg Arg Val Val Gly Ile Asn Ser Lys Ala Val 345 Leu Asp Glu Asn Cys Val Asn Gln Met Asn Tyr Leu Lys Cys Val Val Lys Glu Thr Leu Arg Leu His Pro Pro Leu Pro Leu Leu Ile Ala Arg 375 Glu Thr Ser Ser Ser Val Lys Leu Arg Gly Tyr Asp Ile Pro Ala Lys Thr Met Val Phe Ile Asn Ala Trp Ala Ile Gln Arg Asp Pro Glu Leu 410 Trp Asp Asp Pro Glu Glu Phe Ile Pro Glu Arg Phe Glu Thr Ser Gln 425 Val Asp Leu Asn Gly Gln Asp Phe Gln Leu Ile Pro Phe Gly Ile Gly

291

339

90

									,							
		435					440		-6	_		445				
Arg	Arg 450	Gly	Cys	Pro	Ala	Met 455	Ser	Phe	Gly	Leu	Ala 460	Ser	Thr	Glu	Tyr	
465					470	Tyr				4/5					400	
_				485		Asn			490					400		
Thr	Val	Ser	Lys 500	Lys	Val	Pro	Leu	His 505	Leu	Glu	Pro	Glu	Pro 510	Tyr	LXs	
Thr																
(2)	INF	ORMA'	rion	FOR	SEQ	ID 1	10:3	:								
	(i	() (1	A) LI B) T C) S'	ENGTI YPE : TRANI	H: 10 nuc DEDN	CTERI 691 h leic ESS: line	aci sin	pai d	rs						u.	
	(ii) MO	LECU	LE T	YPE:	CDN	4.									
	(ix) FE. (.	A) N.	AME/	KEY:	CDS	.154	5								
	(xi) SE	QUEN	CE D	ESCR	ITTI	ON:	SEQ	ID N	0:3:						
CCT	AGAT	CTA	TCAT	Me	G GT t Va 1	C AT	G GA t Gl	G CI u Le	T CA u Hi 5	C AA s As	.C CA n Hi	.C AC	I PI	T TT o Ph O	C TC e Se	r 51 r
ATT Ile	TAC	TTC Phe	Ile	ACC Thr	TCC	: ATT	CTC Lev 20	i Pne	T ATT	TTC Phe	TTC Phe	GTG Val		TTC Phe	AAA Lys	99
TTA Leu	GTI Val	Gln	AGA Arg	TCG Ser	GAT Asp	TCC Ser 35	Lys	A ACC	TCC Ser	TCT Ser	TACC Thr	. Cys	AAA Lys	TTG Lev	CCC Pro	147
CCA Pro	Gly	A CCA	AGG Arg	ACA Thr	CTA Leu 50	A CCT 1 Pro	CTO	ATA	A GGC e Gly	AAC ASI 55	J ITE	A CAC	CAG Glr	ATT	GTT Val	
GGC Gl _y	TC/ Sei	A CTO	CCC	GTI Val	. His	TAC	TAC Ty	TTI	A AA u Lys 70	s Ası	r TTC	G GC/	A GAT A Asp	AAC Lys	, 1 y -	243

85

GGT CCA TTA ATG CAT CTA AAA CTA GGA GAG GTG TCC AAC ATC ATA GTC

Gly Pro Leu Met His Leu Lys Leu Gly Glu Val Ser Asn Ile Ile Val

ACT TCC CCA GAA ATG GCC CAA GAG ATT ATG AAG ACA CAT GAT CTC AAC

80

-7-

									-/-	_							
Thr	Ser	Pro 95	Glu	Met	Ala	Gln	Glu 100	Ile	Met	Lys	Thr	His 105	Asp	Leu	Asn		
TTC	TCT	GAT	AGG	CCA	GAC	TTT	GTA	TTG	TCT	AGA	ATA	GTT	TCT	TAC	AAC		387
Phe	Ser 110	Asp	Arg	Pro	Asp	Phe 115	Wal	Leu	Ser	Arg	11e 120	vai	ser	Tyr	ASN		
GGT	TCT	GGC	TTA	GTC	TTC	AGT	CAA	CAT	GGA	GAC	TAT	TGG	AGG	CAA	CTA		435
Gly 125	Ser	Gly	Ile	Val	Phe 130	Ser	Gin	His	GIĀ	135	Tyr	Trp	Arg	Gln	140		
AGA	AAG	ATA	TGC	ACA	GTA	GAG	TTA	CTA	ACA	GCA	AAG	CGC	GIG	CAG	TÇT		483
				145					150					Gln 155			
TTT	CGG	TCC	ATA	AGA	GAA	GAG	GAG	GTG	GCA	GAA	CTA	GTT	AAA	AAA Lys	ATA		531
	_		160					165					170				
GCT	GCA	ACT	GCA	AGT	GAA	GAA	GGG	GGG	TCC	ATT	TTT	AAT	CTC	ACC Thr	CAG Gln		579
		175					180					185					
AGC	ATT	TAC	TCA	ATG	ACT	TTT	GGG Glv	ATA	GCG	GCA Ala	CGA	GCG Ala	GCT Ala	TTT Phe	GGT Glv		627
	190					195					200						
AAA	AAG	AGC	AGA	TAC	CAA	CAA	GTG Val	TTC	ATA Ile	TCA Ser	AAC Asn	ATG Met	CAT His	AAA Lys	CAA Gln		675
205	- rys	261	Arg	171	210	0111				215				•	220		
TTG	ATG	CTT	CTG	GGA	GGG	TTT	TCT	GTT	GCT	GAT	CTC	TAT	CCT	TCT	AGT		723
Leu	Met	Leu	Leu	Gly 225	Gly	Phe	Ser	Val	Ala 230	Asp	Leu	Tyr	Pro	Ser 235	Ser		
AGA	GTG	TTT	CAA	ATG	ATG	GGG	GCG	ACG	GGG	AAA	CTT	GAA	AAA	GTG	CAT		771
Arg	Val	Phe	Gln 240	Met	Met	Gly	Ala	Thr 245	GTÅ	ГÀЗ	Leu	GIU	150	Val	HIS		
AGA	GTG	ACA	GAT	AGG	GTG	TTG	CAA	GAC	ATC	ATC	GAC	GAG	CAC	AAA	AAT		819
Arg	Val	Thr 255	Asp	Arg	Val	Leu	Gln 260	Asp	Ile	Ile	Asp	Glu 265	His	Lys	Asn		
AGA	AAC	AGA	AGC	AGC	GAG	GAG	CGT	GAA	GCA	GTG	GAA	GAT	CTA	GTT	GAT		867
Arg	Asn	Arg	Ser	Ser	Glu	Glu 275	Arg	Glu	Ala	Val	Glu 280	Asp	Leu	Val	Asp		
	270													~			015
GTT	CTT	CTC	AAG	TTT	CAA	AAG Lvs	GAA Glu	TCG Ser	GAA Glu	TTT Phe	CGC	TTG Leu	ACT Thr	GAT Asp	GAC Asp		915
285		БСС	Lys	1110	290	-1-				295				_	300		
AAC	ATT	AAA	GCC	GTC	ATC	CAG	GAC	ATA	TTC	ATT	GGT	GGA	GGC	GAA	ACA		963
Asn	Ile	Lys.	Ala	Val 305		Gln	Asp	Ile	Phe 310	Ile	Gly	GTA	GIÀ	Glu 315	inr		
TCA	TCT	TCT	GTT	GTG	GAA	TGG	GGG	ATG	TCA	GAA	TTG	ATA	AGA	AAC	CCG	;	1011
Ser	Ser	Ser	Val 320		Glu	Trp	Gly	Met 325		Glu	Leu	. ile	Arg 330	Asn	PLO		

									-8	_							
AGG Arg	GTG Val	ATG Met 335	GAA Glu	GAA Glu	GCA Ala	CAA Gln	GCA Ala 340	GAG Glu	GTG Val	AGA Arg	AGA Arg	GTG Val 345	TAT Tyr	GAT Asp	AGC Ser		1059
AAG Lys	GGA Gly 350	TAT Tyr	GTG Val	GAT Asp	GAG Glu	ACA Thr 355	GAA Glu	TTG Leu	CAC His	CAA Gln	TTG Leu 360	ATA Ile	TAC Tyr	TTA Leu	AAG Lys		1107
TCC Ser 365	ATC Ile	ATC	AAA Lys	GAA Glu	ACC Thr 370	ATG Met	AGG Arg	TTA Leu	CAT His	CCA Pro 375	CCT Pro	GTG Val	CCA Pro	TTG Leu	TTA Leu 380		1155
GTT Val	CCT Pro	AGA Arg	GTA Val	AGT Ser 385	AGA Arg	GAA Glu	AGG Arg	TGC Cys	CAA Gln 390	ATC Ile	AAT Asn	GGA Gly	TAT Tyr	GAG Glu 395	ATA Íle		1203
CCC Pro	TCT Ser	AAG Lys	ACT Thr 400	AGG Arg	ATC Ile	ATT Ile	ATC Ile	AAT Asn 405	GCT Ala	TGG Trp	GCA Ala	ATT Ile	GGA Gly 410	AGG Arg	AAT Asn		1251
CCT Pro	AAG Lys	TAT Tyr 415	TGG Trp	GGT Gly	GAA Glu	ACT Thr	GAG Glu 420	AGT Ser	TTT Phe	AAA Lys	CCT Pro	GAG Glu 425	AGG Arg	TTT Phe	CTT Leu		1299
AAT Asn	AGC Ser 430	TCC Ser	ATT Ile	GAT Asp	TTT Phe	AGG Arg 435	GGC Gly	ACA Thr	GAC Asp	TTT Phe	GAA Glu 440	TTT Phe	ATC Ile	CCA Pro	TTT Phe	•	1347
GGT Gly 445	GCT Ala	GGA Gly	AGG Arg	AGG Arg	ATC Ile 450	TGC Cys	CCC Pro	GGC Gly	ATT Ile	ACA Thr 455	Pne	GCC Ala	ATA Ile	CCC Pro	AAC Asn 460		1395
ATT Ile	GAG Glu	TTG Leu	CCA Pro	CTT Leu 465	GCT Ala	CAG Gln	TTA Leu	CTT Leu	TAC Tyr 470	His	TTT	GAT	TGG	AAG Lys 475	CTT Leu	1	1443
CCC Pro	AAT Asn	AAA Lys	ATG Met 480	Lys	AAT Asn	GAA Glu	GAA Glu	CTT Leu 485	GAC Asp	ATG Met	ACG Thr	GAG Glu	TCA Ser 490	AAT Asn	GGA Gly		1491
ATT Ile	ACT Thr	TTA Leu 495	Arg	AGA Arg	CAA Gln	AAT Asn	GAC Asp 500	Leu	TGC Cys	TTG Leu	ATT Ile	CCC Pro 505	ATT Ile	ACT Thr	CGT Arg		1539
	CCT Pro 510		AATG	TAT	GAAC	AATT	AA T	GTCA	TAAA	C TA	TTTA	AGTT	TTA	TCTT	TTA		1595
CTA	CTTC	CAG	CATI	TCGT	'AA T	TGGA	CAAT	G AC	TATO	ATTA	A ACT	TAAG	TTA	CTTC	CTTA	rg	1655
ATT	AACT	TGA	CATA	TGAA	TG A	ACAT	TTCI	A AG	ATAA								1691

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 510 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

-9-

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
- Met Val Met Glu Leu His Asn His Thr Pro Phe Ser Ile Tyr Phe Ile
- Thr Ser Ile Leu Phe Ile Phe Phe Val Phe Phe Lys Leu Val Gln Arg 20 25 30
- Ser Asp Ser Lys Thr Ser Ser Thr Cys Lys Leu Pro Pro Gly Pro Arg
- Thr Leu Pro Leu Ile Gly Asn Ile His Gln Ile Val Gly Ser Leu Pro
- Val His Tyr Tyr Leu Lys Asn Leu Ala Asp Lys Tyr Gly Pro Leu Met 65 70 75 80
- His Leu Lys Leu Gly Glu Val Ser Asn Ile Ile Val Thr Ser Pro Glu 85 90 95
- Met Ala Gln Glu Ile Met Lys Thr His Asp Leu Asn Phe Ser Asp Arg 100 105 110
- Pro Asp Phe Val Leu Ser Arg Ile Val Ser Tyr Asn Gly Ser Gly Ile 115 120 125
- Val Phe Ser Gln His Gly Asp Tyr Trp Arg Gln Leu Arg Lys Ile Cys 130 ·135 140
- Thr Val Glu Leu Leu Thr Ala Lys Arg Val Gln Ser Phe Arg Ser Ile 145 150 155 160
- Arg Glu Glu Val Ala Glu Leu Val Lys Lys Ile Ala Ala Thr Ala 165 170 175
- Ser Glu Glu Gly Gly Ser Ile Phe Asn Leu Thr Gln Ser Ile Tyr Ser 180 185 190
- Met Thr Phe Gly Ile Ala Ala Arg Ala Ala Phe Gly Lys Lys Ser Arg 195 200 205
- Tyr Gln Gln Val Phe Ile Ser Asn Met His Lys Gln Leu Met Leu Leu 210 215 220
- Gly Gly Phe Ser Val Ala Asp Leu Tyr Pro Ser Ser Arg Val Phe Gln 225 230 235
- Met Met Gly Ala Thr Gly Lys Leu Glu Lys Val His Arg Val Thr Asp 245 250 255
- Arg Val Leu Gln Asp Ile Ile Asp Glu His Lys Asn Arg Asn Arg Ser 260 265 270
- Ser Glu Glu Arg Glu Ala Val Glu Asp Leu Val Asp Val Leu Leu Lys 275 280 285
- Phe Gln Lys Glu Ser Glu Phe Arg Leu Thr Asp Asp Asn Ile Lys Ala

-10-

300 295 290 Val Ile Gln Asp Ile Phe Ile Gly Gly Glu Thr Ser Ser Val 315 310 Val Glu Trp Gly Met Ser Glu Leu Ile Arg Asn Pro Arg Val Met Glu Glu Ala Gln Ala Glu Val Arg Arg Val Tyr Asp Ser Lys Gly Tyr Val Asp Glu Thr Glu Leu His Gln Leu Ile Tyr Leu Lys Ser Ile Ile Lys 360 Glu Thr Met Arg Leu His Pro Pro Val Pro Leu Leu Val Pro Arg Val 375 Ser Arg Glu Arg Cys Gln Ile Asn Gly Tyr Glu Ile Pro Ser Lys Thr 395 Arg Ile Ile Asn Ala Trp Ala Ile Gly Arg Asn Pro Lys Tyr Trp 410 Gly Glu Thr Glu Ser Phe Lys Pro Glu Arg Phe Leu Asn Ser Ser Ile Asp Phe Arg Gly Thr Asp Phe Glu Phe Ile Pro Phe Gly Ala Gly Arg Arg Ile Cys Pro Gly Ile Thr Phe Ala Ile Pro Asn Ile Glu Leu Pro 455 Leu Ala Gln Leu Leu Tyr His Phe Asp Trp Lys Leu Pro Asn Lys Met 475 470 Lys Asn Glu Glu Leu Asp Met Thr Glu Ser Asn Gly Ile Thr Leu Arg Arg Gln Asn Asp Leu Cys Leu Ile Pro Ile Thr Arg Leu Pro (2) INFORMATION FOR SEQ ID NO:5: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1644 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 4..1542

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAA ATG GCC ACT CTT TCC TCC TAC GAC CAC TTC ATC TTC ACT GCC TTA

-11-

									-11-	-							
	Met 1	Ala	Thr	Leu	Ser 5	Ser	Tyr	Asp	His	Phe 10	Ile	Phe	Thr	Ala	Leu 15		
GCT Ala	TTC Phe	TTC Phe	ATA Ile	TCT Ser 20	GGC Gly	CTA Leu	ATT Ile	TTC Phe	TTC Phe 25	CTC Leu	AAA Lys	CAG Gln	AAA Lys	TCC Ser 30	AAA Lys		96
TCC Ser	AAA Lys	AAG Lys	TTC Phe 35	AAC Asn	Lea CTC	CCT Pro	CCA Pro	GGA Gly 40	CCC Pro	CCC Pro	GGG Gly	TGG Trp	CCT Pro 45	ATT Ile	GTT Val		144
GGG Gly	AAC Asn	CTC Leu 50	TTC Phe	CAA Gln	GTT Val	GCT Ala	CGT Arg 55	TCT Ser	GGG Gly	AAA Lys	CCT Pro	TTC Phe 60	TTT Phe	GAG Glu	TAT Tyr		192
GTG Val	AAC Asn 65	GAT Asp	GTG Val	AGA Arg	CTC Leu	AAA Lys 70	TAT Tyr	GGC Gly	TCA Ser	ATC Ile	TTC Phe 75	ACC Thr	CTC Leu	AAG Lys	ATG Met		240
GGA Gly 80	ACA Thr	AGG Arg	ACC Thr	ATG Met	ATC Ile 85	ATC Ile	CTC Leu	ACC Thr	GAC Asp	GCA Ala 90	AAA Lys	CTG Leu	GTC Val	CAC His	GAG Glu 95		288
GCC Ala	ATG Met	ATC Ile	CAA Gln	AAG Lys 100	GGT Gly	GCA Ala	ACC Thr	TAC Tyr	GCC Ala 105	ACC Thr	AGG Arg	CCC Pro	CCC Pro	GAG Glu 110	AAC Asn		336
CCC Pro	ACC Thr	AGA Arg	ACC Thr 115	ATC Ile	TTC Phe	AGT Ser	GAA Glu	AAC Asn 120	AAG Lys	TTC Phe	ACC Thr	GTG Val	AAT Asn 125	GCA Ala	GCG Ala	-	384
ACC Thr	TAT Tyr	GGC Gly 130	CCC Pro	GTG Val	TGG Trp	AAG Lys	TCG Ser 135	CTG Leu	AGG Arg	AGG Arg	AAC Asn	ATG Met 140	GTG Val	CAG Gln	AAC Asn		432
ATG Met	CTC Leu 145	Ser	TCA Ser	ACA Thr	AGA Arg	CTT Leu 150	AAG Lys	GAG Glu	TTT Phe	CGC	AGT Ser 155	Val	CGG Arg	GAC Asp	AAT Asn		480
GCG Ala 160	Met	GAC Asp	AAG Lys	CTC Leu	ATC Ile 165	AAC Asn	AGA Arg	CTC Leu	AAG Lys	GAC Asp 170	GAG Glu	GCC Ala	GAG Glu	AAG Lys	AAT Asn 175		528
AAC Asn	GGC Gly	GTG Val	GTT Val	TGG Trp 180	GTG Val	CTC Leu	AAG Lys	GAT Asp	GCC Ala 185	AGG Arg	TTT	GCT Ala	GTT Val	TTT Phe 190	TGC Cys		576
ATA Ile	CTT	GTG Val	GCT Ala 195	Met	TGT Cys	TTT Phe	GGT	CTT Leu 200	Glu	ATG Met	GAT Asp	GAG Glu	GAG Glu 205	ACA Thr	GTG Val		624
GAG Glu	AGA Arg	ATA Ile 210	Asp	CAG Gln	GTT Val	ATG Met	AAG Lys 215	Ser	GTT Val	CTC Leu	ATC Ile	ACT Thr 220	Leu	GAC Asp	CCG Pro		672
AGA Arg	ATT ; Ile 225	Asp	GAC Asp	TAT	CTT Leu	CCA Pro 230	Ile	CTA Leu	AGC Ser	CCC Pro	TTI Phe 235	Phe	TCA Ser	AAG Lys	CAA Gln		720

-12-AGA AAG AAA GCC TTG GAG GTT CGC AGA GAA CAG GTT GAG TTC TTA GTT 768 Arg Lys Lys Ala Leu Glu Val Arg Arg Glu Gln Val Glu Phe Leu Val 245 CCA ATT ATA GAA CAA AGA AGA AGA GCA ATT CAA AAC CCT GGG TCA GAT 816 Pro Ile Ile Glu Gln Arg Arg Ala Ile Gln Asn Pro Gly Ser Asp 265 260 CAC ACC GCC ACA ACG TTT TCC TAC CTA GAC ACA CTT TTT GAC CTC AAA 864 His Thr Ala Thr Thr Phe Ser Tyr Leu Asp Thr Leu Phe Asp Leu Lys 280 275 GTT GAA GGG AAG AAA TCA GCA CCC TCT GAT GCA GAA TTG GTG TCT TTA 912 Val Glu Gly Lys Lys Ser Ala Pro Ser Asp Ala Glu Leu Val Ser Leu 295 TGC TCA GAG TTT CTT AAC GGT GGC ACA GAC ACA ACA GCA ACA GCG GTT 960 Cys Ser Glu Phe Leu Asn Gly Gly Thr Asp Thr Thr Ala Thr Ala Val 310 GAG TGG GGC ATA GCA CAG CTC ATA GCG AAC CCT AAC GTT CAG ACA AAG 1008 Glu Trp Gly Ile Ala Gln Leu Ile Ala Asn Pro Asn Val Gln Thr Lys 330 325 320 CTG TAC GAG GAA ATA AAG AGA ACG GTG GGA GAG AAG AAG GTG GAT GAA 1056 Leu Tyr Glu Glu Ile Lys Arg Thr Val Gly Glu Lys Lys Val Asp Glu AAG GAC GTT GAG AAA ATG CCA TAC CTA CAC GCT GTG GTG AAG GAG CTT 1104 Lys Asp Val Glu Lys Met Pro Tyr Leu His Ala Val Val Lys Glu Leu CTA AGA AAG CAC CCT CCA ACA CAC TTT GTG CTA ACA CAT GCT GTG ACT 1152 Leu Arg Lys His Pro Pro Thr His Phe Val Leu Thr His Ala Val Thr 375 370 GAG CCC ACC ACT TTG GGA GGG TAT GAC ATA CCA ATT GAT GCA AAT GTT 1200 Glu Pro Thr Thr Leu Gly Gly Tyr Asp Ile Pro Ile Asp Ala Asn Val 390 GAG GTG TAC ACA CCA GCC ATT GCT GAG GAC CCC AAA AAT TGG TTA AAC 1248 Glu Val Tyr Thr Pro Ala Ile Ala Glu Asp Pro Lys Asn Trp Leu Asn CCT GAG AAG TTT GAC CCT GAG AGA TTC ATC TCT GGG GGT GAG GAA GCA 1296 Pro Glu Lys Phe Asp Pro Glu Arg Phe Ile Ser Gly Glu Glu Ala 420 GAC ATA ACT GGG GTC ACA GGG GTG AAG ATG ATG CCA TTT GGG GTT GGG 1344 Asp Ile Thr Gly Val Thr Gly Val Lys Met Met Pro Phe Gly Val Gly 440 435 AGA AGG ATT TGC CCT GGC TTG GCT ATG GCC ACA GTG CAT ATT CAC CTC 1392 Arg Arg Ile Cys Pro Gly Leu Ala Met Ala Thr Val His Ile His Leu 455 ATG ATG GCA AGG ATG GTG CAG GAG TTT GAG TGG GGT GCA TAC CCT CCA 1440 Met Met Ala Arg Met Val Gln Glu Phe Glu Trp Gly Ala Tyr Pro Pro 470

									-13-	•						
GAG A Glu A 480	AAG I	AAG Lys	ATG (Asp :	TTC I Phe 1 485	ACT (GGC A	AAG ' Lys '	LLD (GAG Glu 490	TTC . Phe	ACT (GTG Val	GTC Val	ATG Met 495	1488
AAG (GAG '	TCT Ser	Leu .	AGA (ATG .	GCA A	ACC . Thr	ATC . Ile	Lys	CCA A Pro A 505	AGA Arg	GGA Gly	GGA Gly	GAA Glu	AAA Lys 510	GTG Val	1536
AAG Lys		TAAA	ATTT	TC C	TGCT	TCTA	T TC	TTCT	GGGT	TTI	'AAA'	TTC	ACAG	ACAA	CA	1592
TAAA	TATT	'AT I	GCTA	TATT.	C AT	CATC	TATA	ATG	TATA	CAT	CATO	ATGG	TT P	AC		1644
(2)	INFO	RMAT	rion	FOR	SEQ	ID N	0:6:									
			(B)	LEN TYP TOP	GTH: PE: a POLOG	513 minc Y: l	ami aci inea	.no a .d .r	cids	:						
	-		SEQUE						Q ID	NO:	5:					
Met 1												Thr	Ala	Leu 15	Ala -	
Phe	Phe	İle	Ser 20	Gly	Leu	Ile	Phe	Phe 25	Leu	Lys	Gln	Lys	Ser 30	Lys	Ser	
Lys	Lys	Phe 35	Asn	Leu	Pro	Pro	Gly 40	Pro	Pro	Gly	Trp	Pro 45	Ile	Val	Gly	
Asn	Leu 50	Phe	Gln	Val	Ala	Arg 55	Ser	Gly	Lys	Pro	Phe 60	Phe	Glu	Tyr	Val	
Asn 65		Val	Arg	Leu	Lys 70	Tyr	Gly	Ser	Ile	Phe 75	Thr	Leu	Lys	Met	Gly 80	
			Met	85					90		•			93		
			Lys 100					105					110			
		115					120	•				125				
	130		Val			135					140	•				
145	5				150	•				155	•				160	
Met	. Asp	Lys	s Leu	1le 165		Arg	Leu	ı Lys	170	Glu	ı Ala	Glu	Lys	175	a Asn	

-14-

									-14-						
Gly	Val	Val	Trp 180	Val	Leu	Lys	Asp	Ala 185	Arg	Phe	Ala	Val	Phe 190	Cys	Ile
Leu	Val	Ala 195	Met	Cys	Phe	Gly	Leu 200	Glu	Met	Asp	Glu	Glu 205	Thr	Val	Glu
Arg	Ile 210	Asp	Gln	Val	Met	Lys 215	Ser	Val	Leu	Ile	1hr 220	Leu	Asp	Pro	Arg
Ile 225	Asp	Asp	Tyr	Leu	Pro 230	Ile	Leu	Ser	Pro	Phe 235	Phe	Ser	Lys	Gln	Arg 240
Lys	Lys	Ala	Leu	Glu 245	Val	Arg	Arg	Glu	Gln 250	Val	Glu	Phe	Leu	Val 255	Pro
Ile	Ile	Glu	Gln 260	Arg	Arg	Arg	Ala	Ile 265	Gln	Asn	Pro	Gly	Ser 270	Asp	His
Thr	Ala	Thr 275	Thr	Phe	Ser	Tyr	Leu 280	Asp	Thr	Leu	Phe	Asp 285	Leu	Lys	Val
Glu	Gly 290	Lys	Lys	Ser	Ala	Pro 295	Ser	Asp	Ala	Glu	Leu 300	Val	Ser	Leu	Cys
Ser 305	Glu	Phe	Leu	Asn	Gly 310	Gly	Thr	Asp	Thr	Thr 315	Ala	Thr	Ala	Val	Glu 320
Trp	Gly	Ile	Ala	Gln 325	Leu	Ile	Ala	Asn	Pro 330	Asn	Val	Gln	Thr	Lys 335	Leu
Tyr	Glu	Glu	Ile 340	Lys	Arg	Thr	Val	Gly 345	Glu	Lys	Lys	Val	Asp 350	Glu	Lys
Asp	Val	Glu 355	Lys	Met	Pro	Tyr	Leu 360	His	Ala	Val	Val	Lys 365	Glu	Leu	Leu
Arg	Lys 370	His	Pro	Pro	Thr	His 375	Phe	Val	Leu	Thr	His 380	Ala	Val	Thr	Glu
Pro 385	Thr	Thr	Leu	Gly	Gly 390	Tyr	Asp	Ile	Pro	Ile 395	Asp	Ala	Asn	Val	Glu 400
Val	Tyr	Thr	Pro	Ala 405	Ile	Ala	Glu	Asp	Pro 410	Lys	Asn	Trp	Leu	Asn 415	Pro
Glu	Lys	Phe	Asp 420	Pro	Glu	Arg	Phe	Ile 425		Gly	Gly	Glu	Glu 430	Ala	Asp
Ile	Thr	Gly 435	Val	Thr	Gly	Val	Lys 440		Met	Pro	Phe	Gly 445	Val	Gly	Arg
Arg	Ile 450		Pro	Gly	Leu	Ala 455		Ala	Thr	Val	His 460	Ile	His	Leu	Met
Met 465		Arg	Met	Val	Gln 470		Phe	Glu	Trp	Gly 475	Ala	Tyr	Pro	Pro	Glu 480
Lys	Lys	Met	Asp	Phe 485		Gly	Lys	Trp	Glu 490	Phe	Thr	Val	Val	Met 495	Lys

-15-

Glu Ser Leu Arg Ala Thr Ile Lys Pro Arg Gly Gly Glu Lys Val Lys 500 505 510

Leu

- (2) INFORMATION FOR SEQ ID NO:7:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1611 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 20..1588
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AAGCACTATC CCTCCCACC	ATG Met	Thr	Ser	His	Ile	Asp	Asp	Asn	Leu	TGG Trp	ATA Ile	52
	1							~~» ′		מיים (CTT	100

ATA GCC CTG ACC TCG AAA TGC ACC CAA GAA AAC CTT GCA TGG GTC CTT 10

Ile Ala Leu Thr Ser Lys Cys Thr Gln Glu Asn Leu Ala Trp Val Leu

15 20 25

TTG ATC ATG GGC TCA CTC TGG TTA ACC ATG ACT TTC TAT TAC TGG TCA

Leu Ile Met Gly Ser Leu Trp Leu Thr Met Thr Phe Tyr Tyr Trp Ser

30 35 40

CAC CCC GGT GGT CCT GCC TGG GGC AAG TAC TAC ACC TAC TCT CCC CCC

His Pro Gly Gly Pro Ala Trp Gly Lys Tyr Tyr Thr Tyr Ser Pro Pro

45 50 55

CTT TCA ATC ATT CCC GGT CCC AAA GGC TTC CCT CTT ATT GGA AGC ATG
Leu Ser Ile Ile Pro Gly Pro Lys Gly Phe Pro Leu Ile Gly Ser Met
60 65 70 75

GGC CTC ATG ACT TCC CTG GCC CAT CAC CGT ATC GCA GCC GCG GCC GCC Gly Leu Met Thr Ser Leu Ala His His Arg Ile Ala Ala Ala Ala 80 85 90

ACA TGC AGA GCC AAG CGC CTC ATG GCC TTT AGT CTC GGC GAC ACA CGT

Thr Cys Arg Ala Lys Arg Leu Met Ala Phe Ser Leu Gly Asp Thr Arg

95 100 105

GTC ATC GTC ACG TGC CAC CCC GAC GTG GCC AAG GAG ATT CTC AAC AGC
Val Ile Val Thr Cys His Pro Asp Val Ala Lys Glu Ile Leu Asn Ser
110 115 120

TCC GTC TTC GCC GAT CGT CCC GTC AAA GAA TCC GCA TAC AGC CTC ATG

Ser Val Phe Ala Asp Arg Pro Val Lys Glu Ser Ala Tyr Ser Leu Met

125

130

135

-16-

TTT Phe 140	AAC Asn	CGC Arg	GCC Ala	ATC Ile	GGC Gly 145	TTC Phe	GCC Ala	TCT Ser	TAC Tyr	GGA Gly 150	GTT Val	TAC Tyr	TGG Trp	CGA Arg	AGC Ser 155		484
CTC Leu	AGG Arg	AGA Arg	ATC Ile	GCC Ala 160	TCT Ser	AAT Asn	CAC His	CTC Leu	TTC Phe 165	TGC Cys	CCC Pro	CGC Arg	CAG Gln	ATA Ile 170	AAA Lys		532
GCC Ala	TCT Ser	GF.G Glu	CTC Leu 175	CAA Gln	CGC	TCT Ser	CAA Gln	ATC Ile 180	GCC Ala	GCC Ala	CAA Gln	ATG Met	GTT Val 185	CAC His	ATC Ile		580
CTA Leu	AAT Asn	AAC Asn 190	AAG Lys	CGC Arg	CAC His	CGC Arg	AGC Ser 195	TTA Leu	CGT Arg	GTT Val	CGC Arg	CAA Gln 200	GTG Val	CTG Leu	ÁAA Lys		628
AAG Lys	GCT Ala 205	TCG Ser	CTC Leu	AGT Ser	AAC Asn	ATG Met 210	ATG Met	TGC Cys	TCC Ser	GTG Val	TTT Phe 215	GGA Gly	CAA Gln	GAG Glu	TAT		676
AAG Lys 220	CTG Leu	CAC His	GAC Asp	CCA Pro	AAC Asn 225	AGC Ser	GGA Gly	ATG Met	GAA Glu	GAC Asp 230	CTT Leu	GGA Gly	ATA Ile	TTA Leu	GTG Val 235		724
GAC Asp	CAA Gln	GGT Gly	TAT Tyr	GAC Asp 240	CTG Leu	TTG Leu	GGC Gly	CTG Leu	TTT Phe 245	AAT Asn	TGG Trp	GCC Ala	GAC Asp	CAC His 250	CTT Leu		772
CCT Pro	TTT Phe	CTT Leu	GCA Ala 255	CAT His	TTC Phe	GAC Asp	GCC Ala	CAA Gln 260	Asn	ATC Ile	CGG Arg	TTC Phe	AGG Arg 265	TGC Cys	TCC Ser		820
Asn	Leu	Val 270	Pro	Met	Val	Asn	Arg 275	Phe	Val	Gly	Thr	280	iie	Ala	GAA Glu	•	868
His	Arg 285	Ala	Ser	Lys	Thr	Glu 290	Thr	Asn	Arg	Asp	Phe 295	Val	Asp	vaı	TTG Leu		916
Leu 300	Ser	Leu	Pro	Glu	Pro 305	Asp	Gln	Leu	Ser	310	Ser	Asp	Met	116	GCT Ala 315		964
Val	Leu	Trp	Glu	Met 320	Ile	Phe	Arg	Gly	325	Asp	Thr	· vai	Ala	330			1012
ATA Ile	GAG Glu	TGG	ATA Ile 335	Leu	GCG Ala	AGG Arg	ATG Met	GCG Ala 340	Leu	CAT His	CCI Pro	CAT His	GTG Val 345	GID	TCC Ser		1060
Lys	Val	350	Glu	. Glu	Leu	Asp	355	. Val	. Val	. Gly	r Lys	360	. Arg	A1a	GTC Val		1108
GCA Ala	GAG Glu 365	. Asp	GAC Asp	GTG Val	GCA Ala	GTG Val 370	. Met	ACC Thr	TAC	CTA	CCA Pro 375) Ala	GTG Val	GTC Val	AAG Lys		1156

-17-

GAG Glu 380	GTG Val·	CTG Leu	CGG Arg	CTG Leu	CAC His 385	CCG Pro	CCG Pro	GGC Gly	CCA Pro	CTT Leu 390	CTA Leu	TCA Ser	TGG Trp	GCC Ala	CGC Arg 395	1204
TTG Leu	TCC Ser	ATC Ile	Asn	GAT Asp 400	ACG Thr	ACC Thr	ATT Ile	GAT Asp	GGG Gly 405	TAT Tyr	CAC His	GTA Val	CCT Pro	GCG Ala 410	GGG Gly	1252
ACC Thr	ACT Thr	GCT Ala	ATG Met 415	GTC Val	AAC Asn	ACG Thr	TGG Trp	GCT Ala 420	ATT Ile	TGC Cys	AGG Arg	GAC Asp	CCA Pro 425	CAC His	GTG Val	1300
TGG Trp	AAG Lys	GAC Asp 430	CCA Pro	CTC Leu	GAA Glu	TTT Phe	ATG Met 435	CCC Pro	GAG Glu	AGG Arg	TTT Phe	GTC Val 440	ACT Thr	GCG Ala	GGT Gly	1348
GGA Gly	GAT Asp 445	GCC Ala	GAA Glu	TTT Phe	TCG Ser	ATA Ile 450	CTC Leu	GGG Gly	TCG Ser	GAT Asp	CCA Pro 455	AGA Arg	CTT Leu	GCT Ala	CCA Pro	1396
TTT Phe 460	GGG Gly	TCG Ser	GGT Gly	AGG Arg	AGA Arg 465	GCG Ala	TGC Cys	CCA Pro	GGG Gly	AAG Lys 470	ACT Thr	CTT Leu	GGA Gly	TGG Trp	GCT Ala 475	1444
ACG Thr	GTG Val	AAC Asn	TTT Phe	TGG Trp 480	GTG Val	GCG Ala	TCG Ser	CTC Leu	TTG Leu 485	CAT	GAG Glu	TTC Phe	GAA Glu	TGG Trp 490	GTA Val	1492
CCG Pro	TCT Ser	GAT Asp	GAG Glu 495	AAG Lys	GGT Gly	GTT Val	GAT Asp	CTG Leu 500	Thr	GAG Glu	GTG Val	CTG Leu	AAG Lys 505	CTC Leu	TCT Ser	1540
AGT Ser	GAA Glu	ATG Met 510	GCT Ala	AAC Asn	CCT Pro	CTC Leu	ACC Thr 515	Val	AAA Lys	GTG Val	CGC Arg	CCC Pro 520	Arg	CGT Arg	GGA Gly	1588
TAA	GAGA	GAG	TTGA	AGCT	TT T	ΑT										1611

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 523 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Thr Ser His Ile Asp Asp Asn Leu Trp Ile Ile Ala Leu Thr Ser

Lys Cys Thr Gln Glu Asn Leu Ala Trp Val Leu Leu Ile Met Gly Ser

Leu Trp Leu Thr Met Thr Phe Tyr Tyr Trp Ser His Pro Gly Gly Pro 35 40 45

Ala	Trp 50	Gly	Lys	Tyr	Tyr	Thr 55	Tyr	Ser	Pro	Pro	Leu 60	Ser	Ile	Ile	Pro
Gly 65	Pro	Lys	Gly	Phe	Pro 70	Leu	Ile	Gly	Ser	Met 75	Gly	Leu	Met	Thr	Ser 80
Leu	Ala	His	His	Arg 85	Ile	Ala	Ala	Ala	Ala 90	Ala	Thr	Cys	Arg	Ala 95	Lys
Arg	Leu	Met	Ala 100	Phe	Ser	Leu	Gly	Asp 105	Thr	Arg	Val	Ile	Val 110	Ihr	Cys
His	Pro	Asp 115	Val	Ala	Lys	Glu	Ile 120	Leu	Asn	Ser	Ser	Val 125	Phe	Ala	Asp
Arg	Pro 130	Val	Lys	Glu	Ser	Ala 135	Tyr	Ser	Leu	Met	Phe 140	Asn	Arg	Ala	Ile
Gly 145	Phe	Ala	Ser	Tyr	Gly 150	Val	Tyr	Trp	Arg	Ser 155	Leu	Arg	Arg	Ile	Ala 160
				165					170					Leu 175	
			180					185					100	Lys	
		195					200					203		Leu	
	210					215					220			Asp	
Asn 225		Gly	. Met	Glu	Asp 230	Leu	Gly	Ile	Leu	Val 235	Asp	Gln	Gly	Tyr	Asp 240
Leu	Leu	Gly	Leu	Phe 245		Trp	Ala	Asp	His 250	Leu	Pro	Phe	Leu	Ala 255	His
			260)				265	•				270		
		275	5				280					263	•	. Ser	
Thi	Glu 290		. Asr	n Arg	Asp	295	val	. Asp	val	. Leu	1 Leu 300	Ser	Leu	Pro	Glu
Pro 309		Glr	ı Lev	ı Ser	310	Ser	Asp) Met	: Ile	315	a Val	. Lev	ı Trp	Glu	320
				325	5				33(J				339	
			340	0				343	5				330		
Le	u As	p Ala 35		l Vai	l Gly	y Lys	360	a Arg	g Ala	a Va	l Ala	36!	a Asp) Ası	Va]

									-19-							
Ala	Val 370	Met	Thr	Tyr	Leu	Pro 375	Ala	Val	Val	Lys	Glu 380	Val	Leu	Arg	Leu	
His 385	Pro	Pro	Gly	Pro	Leu 390	Leu	3er	Trp	Ala	Arg 395	Leu	Ser	Ile	Asn	Asp 400	
Thr	Thr	Ile	Asp	Cly 405	Tyr	His	Val	Pro	Ala 410	Gly	Thr	Thr	Ala	Met 415	Val	
Asn	Thr	Trp	Ala 420	Ile	Cys	Arg	Asp	Pro 425	His	Val	Trp	Lys	Asp 430	Pro	Leu	
Glu	Phe	Met 435	Pro	Glu	Arg	Phe	Val 440	Thr	Ala	Gly	Gly	Asp 445	Ala	Glu	Þhe	
Ser	Ile 450	Leu	Gly	Ser	Asp	Pro 455	Arg	Leu	Ala	Pro	Phe 460	Gly	Ser	Gly	Arg	
Arg 465	Ala	Cys	Pro	Gly	Lys 470	Thr	Leu	Gly	Trp	Ala 475	Thr	Val	Asn	Phe	Trp 480	
Val	Ala	Ser	Leu	Leu 485	His	Glu	Phe	Glu	Trp 490	Val	Pro	Ser	Asp	Glu 495	Lys	
Gly	Val	Asp	Leu 500	Thr	Glu	Val	Leu	Lys 505	Leu	Ser	Ser	Glu	Met 510	Ala	Asn	
Pro	Leu	Thr 515		Lys	Val	Arg	Pro 520	Arg	Arg	Gly						
(2)	INF	ORMA	TION	FOR	SEQ	ID 1	9: 07	:								
	(i	() () ()	QUEN A) L B) T C) S D) T	ENGT: YPE: TRAN: OPOL	H: 1 nuc DEDN OGY:	788 l leic ESS: lin	acio sing	pai: d	rs							
	(ii) MO	LECU	LE T	YPE:	CDM	A									
	(ix		ATUR A) N B) L	AME/			1601									
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:9:						
GGG	TC A	TG G et G	GC A	TG G et A	CC A la M	TG G et A 5	AT G sp A	CT T la P	TC C	AG C ln H	AC C is G 10	AA A ln T	CT C hr L	TC A eu I	TT le	47
TCC Ser	Ile	- : ATT	CTG	GCC Ala	ATG Met	Leu	GTA Val	. GGC . Gly	GTG	TTG Leu 25	Ile	TAT Tyr	GGC Gly	TTA Leu	AAG Lys 30	95
AGA Arg	ACA Thr	CAT His	AGT Ser	GGC Gly	His	GGC	AAG	ATC	TGT Cys	Ser	GCA Ala	CCT	CAA Gln	GCA Ala 45	GGA Gly	143

-20-

GG. G1	A GCA y Ala	TGG Trp	CCA Pro 50	ATT Ile	ATT Ile	GGC Gly	CAT His	TTA Leu 55	CAC His	CTC Leu	TTT Phe	GGG Gly	GGT Gly 60	CAT His	CAA Gln		191
CA' Hi	T ACT s Thr	CAC His	AAA Lys	ACA Thr	CTT Leu	GGG Gly	ATA Ile 70	ATG Met	GCA. Ala	GAG Glu	AAA Lys	CAT His 75	GGA Gly	CCA Pro	ATT Ile		239
TT Ph	C ACA e Thr 80	Ile	AAG Lys	CTT Leu	GGT Gly	TCA Ser 85	TAC Tyr	AAA Lys	GTT Val	CTT Leu	GTA Val 90	TTG Leu	AGT Ser	AGC Ser	TGG Trp		287
GA Gl 9	G ATG u Met 5	GCC Ala	AAG Lys	GAG Glu	TGT Cys 100	TTC Phe	ACT Thr	GTC Val	CAT His	GAC Asp 105	AAA Lys	GCA Ala	TTT Phe	TCT Ser	ACC Thr 110		335
AG Ar	A CCC g Pro	TGT Cys	GTT Val	GCA Ala 115	GCC Ala	TCA Ser	AAG Lys	CTA Leu	ATG Met 120	GGC Gly	TAC Tyr	AAC Asn	TAT Tyr	GCC Ala 125	ATG Met		383
TT Ph	T GGC e Gly	TTC Phe	ACT Thr 130	CCT Pro	TAT Tyr	GGT Gly	CCT Pro	TAT Tyr 135	TGG Trp	CGT Arg	GAG Glu	ATA Ile	AGG Arg 140	AAA Lys	TTA Leu		431
AC Th	T ACT r Thr	ATT Ile 145	Gln	CTT Leu	CTA Leu	TCT Ser	AAC Asn 150	CAC His	CGG Arg	CTT Leu	GAA Glu	CTG Leu 155	CTG Leu	AAG Lys	AAC Asn		479
AC Th	A AGA r Arg 160	Thr	TCT	GAG Glu	TCA Ser	GAA Glu 165	GTT Val	GCA Ala	ATA Ile	AGA Arg	GAG Glu 170	CTT	TAT Tyr	AAG Lys	TTG Leu		527
TG Tr 17	G TCT p Ser 5	AGA Arg	GAA Glu	GGT Gly	TGT Cys 180	CCA Pro	AAG Lys	GGA Gly	GGG Gly	GTT Val 185	TTG Leu	GTA Val	GAT Asp	ATG Met	AAG Lys 190		575
CA Gl	G TGG n Trp	TTT	GGG Gly	GAT Asp 195	TTA Leu	ACT Thr	CAT His	AAT Asn	ATT Ile 200	GTT Val	CTG Leu	AGA Arg	ATG Met	GTG Val 205	AGA Arg	•	623
G1	G AAG y Lys	CCA Pro	TAC Tyr 210	TAT	GAT Asp	GGT Gly	GCT Ala	AGT Ser 215	GAT Asp	GAT Asp	TAT Tyr	GCA Ala	GAA Glu 220	GGT Gly	GAA Glu		671
GC Al	A AGA a Arg	AGG Arg 225	Tyr	AAG Lys	AAA Lys	GTT Val	ATG Met 230	Gly	GAG Glu	TGT Cys	GTG Val	AGT Ser 235	TTG Leu	TTT	GGG Gly		719
GI Va	G TTI l Phe 240	Val	TTA Leu	TCT Ser	GAT Asp	GCT Ala 245	ATT	CCA Pro	TTT	CTG Leu	GGG Gly 250	TGG Trp	TTG Leu	GAC Asp	ATC		767
AA As 25	C GGA in Gly	TAT Tyr	GAA Glu	AAG Lys	GCC Ala 260	Met	AAG Lys	AGA Arg	ACT Thr	GCA Ala 265	Ser	GAA Glu	TTG Leu	GAT Asp	CCT Pro 270		815
CT Le	G GTT	GAA Glu	. GGG . Gly	TGG Trp	TTA Leu	GAG Glu	GAA Glu	CAC His	AAA Lys	AGG Arg	AAA Lys	AGA Arg	GCT Ala	TTC Phe	AAT Asn		863

			-21-	
	275		280	285
ATG GAT GCA AA Met Asp Ala Ly 29	s Glu Glu (AG GAT AAT Sln Asp Asn 295	TTC ATG GAT GTC ATG Phe Met Asp Val Me 30	t Leu Asn
GTT CTG AAA GA Val Leu Lys As 305	AT GCA GAG A sp Ala Glu I	ATT TCT GGT (le Ser Gly 310	TAT GAT TCA GAT AC Tyr Asp Ser Asp Th 315	C ATC ATC 959 r Ile Ile
AAG GCT ACT TO Lys Ala Thr Cy 320	ys Leu Asn I	CTG ATT TTA Leu Ile Leu 325	GCA GGA AGC GAC AC Ala Gly Ser Asp Th 330	C ACC ATG 1007 r Thr Met
ATT TCA CTA AC Ile Ser Leu Th 335	CA TGG GTG (hr Trp Val 1 340	CTA TCT CTG Leu Ser Leu	CTA CTT AAC CAT CA Leu Leu Asn His Gl 345	A ATG GAA 1055 n Met Glu 350
CTA AAA AAA G Leu Lys Lys Va	TC CAA GAT (al Gln Asp (355	BAA TTG GAC Blu Leu Asp	ACT TAT ATT GGG AA Thr Tyr Ile Gly Ly 360	G GAC AGG 1103 s Asp Arg 365
Lys Val Glu G	AA TCT GAC lu Ser Asp 70	ATA ACC AAG Ile Thr Lys 3 7 5	TTG GTG TAC CTC CA Leu Val Tyr Leu Gl 38	n Ala ile
GTG AAG GAA AG Val Lys Glu Ti 385	CA ATG CGG hr Met Arg	CTG TAT CCA Leu Tyr Pro 390	CCA AGT CCT CTT AT Pro Ser Pro Leu Il 395	C ACC CTT 1199 e Thr Leu
CGT GCA GCC A Arg Ala Ala M 400	TG GAA GAC et Glu Asp	TGC ACC TTC Cys Thr Phe 405	TCA GGT GGC TAT CA Ser Gly Gly Tyr Hi 410	AC ATT CCT 1247
GCT GGG ACA C Ala Gly Thr A 415	GT TTA ATG rg Leu Met 420	GTG AAT GCT Val Asn Ala	TGG AAG ATC CAC CO Trp Lys Ile His Ar 425	GG GAT GGT 1295 Fg Asp Gly 430
CGT GTT TGG A Arg Val Trp S	GT GAT CCT er Asp Pro 435	CAT GAT TTC His Asp Phe	AAG CCT GGA AGG TT Lys Pro Gly Arg Pb 440	TC TTG ACA 1343 ne Leu Thr 445
Ser His Lys A	AT GTT GAT sp Val Asp	GTG AAG GGT Val Lys Gly 455	CAG AAC TAT GAG CT Gln Asn Tyr Glu Le	eu Val Pro
TTT GGT TCT G Phe Gly Ser G 465	GA AGG AGA Bly Arg Arg	GCA TGC CCT Ala Cys Pro 470	GGA GCC TCG CTG GC Gly Ala Ser Leu Al 475	CT CTG CGT 1439 La Leu Arg
GTG GTG CAC T Val Val His L 480	TTG ACC ATG Leu Thr Met	GCT AGA CTO Ala Arg Leu 485	TTA CAT TCT TTC A Leu His Ser Phe A 490	AT GTT GCT 1487 sn Val Ala
TCT CCT TCA A Ser Pro Ser A 495	AAT CAA GTT Asn Gln Val 500	GTG GAC ATO Val Asp Met	G ACA GAG AGC ATT G Thr Glu Ser Ile G 505	GA CTC ACA 1535 ly Leu Thr 510
AAT TTA AAA G	CA ACC CCG	CTT GAA AT	CTC CTA ACT CCA C	GT CTA GAC 1583

									-22-	-							
Asn i	Leu :	Lys .	Ala '	Thr 515	Pro :	Leu (Glu	Ile	Leu 1 520	Leu '	Thr 1	Pro A	Arg :	Leu 525	Asp		
ACC . Thr	AAP. Lys	Leu	TAT (Tyr (GAG . Glu .	AAC Asn	TAGA'	AATT	AT T	AAGC'	TAGT	T TT	CTCC	CAAA			16:	31
TAAG	GGGA	.GG G	GTCC	TCTA	G GT	CCTG	AAAT	CGG	GTAA	TAA	CAAT	AACA	TG G	TTAA	TGCAG	3 16	91
CTTC	CATG	TA G	GATA	ATGA	T TA	TTCA	CTCA	. TGG	GTCA	CCT	TTTA	ATGG	AG C	CTCA	CTGTA	17	51
TTAT	AATA	AC T	'CCAA	ACIT	G TG	GGTC	ACAA	TCC	CCCC							17	88
(2)			CION	NCE	CHAR	ACTE	RIST	CS:	-: d-								
			(B)	TYP	GTH: E: a OLOG	mino	aci	.d	Clus								
	(i	.i) M	OLEC	ULE	TYPE	: pr	otei	.n									
	(×	ai) S	EQUE	NCE	DESC	RIPT	'ION:	SEQ	ID	ио : 1	.0:				ė		
Met 1	Gly	Met	Ala	Met 5	Asp	Ala	Phe	Gln	His 10	Gln	Thr	Leu	Ile	Ser 15	Ile		
Ile	Leu	Ala	Met 20	Leu	Val	Gly	Val	Leu 25	Ile	Tyr	Gly	Leu	Lys 30	Arg	Thr		
His	Ser	Gly 35	His	Gly	Lys	Ile	Cys 40	Ser	Ala	Pro	Gln	Ala 45	Gly	Gly	Ala	•	
Trp	Pro 50	Ile	Ile	Gly	His	Leu 55	His	Leu	Phe	Gly	Gly 60	His	Gln	His	Thr		
His 65	Lys	Thr	Leu	Gly	Ile 70	Met	Ala	Glu	Lys	His 75	Gly	Pro	Ile	Phe	Thr 80		
Ile	Lys	Leu	Gly	Ser 85	Tyr	Lys	Val	Leu	Val 90	Leu	Ser	Ser	Trp	Glu 95	Met		
Ala	Lys	Glu	Cys 100	Phe	Thr	Val	His	Asp 105	Lys	Ala	Phe	Ser	Thr 110	Arg	Pro		
Cys	Val	Ala 115	Ala	Ser	Lys	Leu	Met 120	Gly	Tyr	Asn	Tyr	Ala 125	Met	Phe	Gly		
Phe	Thr		Tyr	Gly	Pro	Tyr 135	Trp	Arg	Glu	Ile	Arg 140	Lys	Leu	Thr	Thr		
Ile 145		Leu	Leu	Ser	Asn 150		Arg	Leu	Glu	Leu 155	Leu	Lys	Asn	Thr	Arg 160		
Thr	Ser	Glu	Ser	Glu 165		Ala	Ile	Arg	Glu 170	Leu	Tyr	Lys	Leu	Trp 175	Ser		
Arg	Glu	Gly	Cys		Lys	Gly	Gly	Val 185	Leu	Val	Asp	Met	Lys 190	Gln	Trp		

							•								
Phe	Gly	Asp 195	Leu	Thr	His	Asn	Ile 200	Val	Leu	Arg	Met	Val 205	Arg	Gly	Lys
Pro	Tyr 210	Tyr	Asp	Gly	Ala	Ser 215	Asp	Asp	Tyr	Ala	Glu 220	Gly	Glu	Ala	Arg
Arg 225	Tyr	Lys	Lys	Val	Met 230	Gly	Glu	Cys	Val	Ser 235	Leu	Phe	Gly	Val	Phe 240
Val	Leu	Ser	Asp	Ala 245	Ile	Pro	Phe	Leu	Gly 250	Trp	Leu	Asp	Ile	Asn 255	Gly
Tyr	Glu	Lys	Ala 260	Met	Lys	Arg	Thr	Ala 265	Ser	Glu	Leu	Asp	Pro 270	Leu	Vál
Glu	Gly	Trp 275	Leu	Glu	Glu	His	Lys 280	Arg	Lys	Arg	Ala	Phe 285	Asn	Met	Asp
Ala	Lys 290	Glu	Glu	Gln	Asp	Asn 295	Phe	Met	Asp	Val	Met 300	Leu	Asn	Val	Leu
Lys 305	Asp	Ala	Glu	Ile	Ser 310	Gly	Tyr	Asp	Ser	Asp 315	Thr	Ile	Ile	Lys	Ala 320
Thr	Cys	Leu	Asn	Leu 325	Ile	Leu	Ala	Gly	Ser 330	Asp	Thr	Thr	Met	11e 335	Ser
Leu	Thr	Trp	Val 340	Leu	Ser	Leu	Leu	Leu 345	Asn	His	Gln	Met	Glu 350	Leu	Lys
Lys	Val	Gln 355	Asp	Glu	Leu	Asp	Thr 360	Tyr	Ile	Gly	Lys	Asp 365	Arg	Lys	Val
Glu	Glu 370	Ser	Asp	Ile	Thr	Lys 375	Leu	Val	Tyr	Leu	Gln 380	Ala	Ile	Val	Lys
Glu 385	Thr	Met	Arg	Leu	Tyr 390	Pro	Pro	Ser	Pro	Leu 395	Ile	Thr	Leu	Arg	Ala 400
Ala	Met	Glu		Cys 405				Gly					Pro	Ala 415	Gly
Thr	Arg	Leu	Met 420	Val	Asn	Ala	Trp	Lys 425	Ile	His	Arg	Asp	Gly 430	Arg	Val
Trp	Ser	Asp 435	Pro	His	Asp	Phe	Lys 440	Pro	Gly	Arg	Phe	Leu 445	Thr	Ser	His
Lys	Asp 450		Asp	Val	Lys	Gly 455		Asn	Tyr	Glu	Leu 460	Val	Pro	Phe	Gly
Ser 465	Gly	Arg	Arg	Ala	Cys 470		Gly	Ala	Ser	Leu 475		Leu	Arg	Val	Val 480
His	Leu	Thr	Met	Ala 485		Leu	Leu	His	Ser 490	Phe	Asn	Val	Ala	Ser 495	Pro
Ser	Asn	Gln	Val 500		Asp	Met	Thr	Glu 505		Ile	Gly	Leu	Thr 510	Asn	Leu

-24-

Lys Ala Thr Pro Leu Glu Ile Leu Leu Thr Pro Arg Leu Asp Thr Lys 515 520 525

Leu Tyr Glu Asn 530

- (2) INFORMATION FOR SEQ ID NO:11:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1657 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TCPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..1548
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CTT Leu 1	GTT Val	CTT Leu	CTT Leu	TCT Ser 5	CTA Leu	TTG Leu	TCT Ser	ATA Ile	GTC Val 10	ATC Ile	TCC Ser	ATT Ile	GTT Val	CTC Leu 15	TTC Phe	48	
ATT Ile	ACC Thr	CAC His	ACA Thr 20	CAC His	AAA Lys	AGA Arg	AAC Asn	AAC Asn 25	ACT Thr	CCA Pro	AGA Arg	GGA Gly	CCA Pro 30	CCA Pro	GGT Gly	96	
CCT Pro	CCA Pro	CCT Pro 35	CTT Leu	CCT Pro	CTC Leu	ATC Ile	GGC Gly 40	AAC Asn	CTT Leu	CAC His	CAA Gln	CTC Leu 45	CAC His	AAC Asn	TCA Ser	144	
TCC Ser	CCA Pro 50	CAT His	CTC Leu	TGC Cys	CTA Leu	TGG Trp 55	CAA Gln	CTC Leu	GCC Ala	AAA Lys	CTC Leu 60	CAC His	GGT Gly	CCT Pro	CTC Leu	192	
ATG Met 65	TCG Ser	TTT Phe	CGC Arg	CTC Leu	GGC Gly 70	GCC Ala	GTG Val	CAA Gln	ACC Thr	GTC Val 75	GTG Val	GTT Val	TCA Ser	TCG Ser	GCC Ala 80	240	
AGA Arg	ATC Ile	GCC Ala	GAA Glu	CAA Gln 85	ATC Ile	TTG Leu	AAA Lys	ACC Thr	CAC His 90	GAC Asp	CTC Leu	AAC Asn	TTC Phe	GCT Ala 95	TCC Ser	288	
AGG Arg	CCT Pro	CTC Leu	TTC Phe 100	GTG Val	GGC Gly	CCG Pro	AGA Arg	AAG Lys 105	CTC Leu	TCT Ser	TAC	GAC Asp	GGG Gly 110	TTG Leu	GAC Asp	336	
ATG Met	GGC Gly	TTC Phe 115	GCA Ala	CCG Pro	TAC Tyr	GGC Gly	CCG Pro 120	TAC Tyr	TGG Trp	AGA Arg	GAA Glu	ATG Met 125	AAG Lys	AAA Lys	CTC Leu	384	
TGC	ATC	GTT	CAC	CTC	TTC	AGC	GCG	CAA	CGC	GTT	CGG	TCC	TTT	CGA	CCA	432	

Cys Ile Val His Leu Phe Ser Ala Gln Arg Val Arg Ser Phe Arg Pro

									-25	5 –						
	130					135					140					
ATT Ile 145	CGA Arg	GAG Glu	AAC Asn	GAG Glu	GTT Val 150	GCA Ala	AAA Lys	ATG Met	GTT Val	CGG Arg 155	AAA Lys	CTG Leu	TCG Ser	GAA Glu	CAC His 160	480
GAA Glu	GCT Ala	TCG Ser	GGT Gly	AC1 Thr 165	GTC Val	GTG Val	AAC Asn	TTG Leu	ACC Thr 170	GAA Glu	ACT Thr	TTG Leu	ATG Met	TCT Ser 175	TTC Phe	528
ACG Thr	AAC Asn	TCT Ser	TTG Leu 180	ATA Ile	TGC Cys	AGA Arg	ATC Ile	GCG Ala 185	TTG Leu	GGG Gly	AAA Lys	AGT Ser	TAC Tyr 190	GGT Gly	TGT Cys	576
GAG Glu	TAC Tyr	GAG Glu 195	GAA Glu	GTA Val	GTT Val	GTT Val	GAT Asp 200	GAG Glu	GTA Val	CTG Leu	GGA Gly	AAC Asn 205	CGG Arg	AGG Arg	AGC Ser	624
AGG Arg	TTG Leu 210	CAG Gln	GTT Val	CTG Leu	CTC Leu	AAC Asn 215	GAG Glu	GCT Ala	CAA Gln	GCG Ala	TTG Leu 220	CTT Leu	TCG Ser	GAG Glu	TTT Phe	672
TTC Phe 225	TTT Phe	TCG Ser	GAT Asp	TAT Tyr	TTT Phe 230	CCG Pro	CCT Pro	ATA Ile	GGA Gly	AAG Lys 235	TGG Trp	GTT Val	GAT Asp	AGA Arg	GTG Val 240	720
ACG Thr	GGA Gly	ATT Ile	CTA Leu	TCG Ser 245	CGG Arg	CTT Leu	GAT Asp	AAA Lys	ACG Thr 250	TTC Phe	AAG Lys	GAG Glu	TTG Leu	GAC Asp 255	GCG Ala	768
TGC Cys	TAC Tyr	GAA Glu	CGA Arg 260	TCA Ser	TCC Ser	TAT Tyr	GAT Asp	CAC His 265	ATG Met	GAT Asp	TCG Ser	GCA Ala	AAG Lys 270	AGT Ser	GGT	816
AAA Lys	AAA Lys	GAT Asp 275	AAT Asn	GAC Asp	AAC Asn	AAA Lys	GAA Glu 280	GTC Val	AAA Lys	GAT Asp	ATT Ile	ATT Ile 285	GAT Asp	ATT Ile	CTT Leu	864
CTC Leu	CAG Gln 290	CTA Leu	Leu	Asp	GAT Asp	Arg	Ser	TTC Phe	ACC Thr	TTT Phe	GAT Asp 300	CTC Leu	ACT Thr	CTC Leu	GAC Asp	912
CAC His 305	ATA Ile	AAA Lys	GCC Ala	GTG Val	CTC Leu 310	ATG Met	AAC Asn	ATC Ile	TTT Phe	ATA Ile 315	GCA Ala	GGA Gly	ACA Thr	GAC Asp	CCG Pro 320	960
AGT Ser	TCC	GCG Ala	ACA Thr	ATA Ile 325	GTT Val	TGG Trp	GCA Ala	ATG Met	AAT Asn 330	GCA Ala	CTG Leu	TTG Leu	AAG Lys	AAT Asn 335	CCC Pro	1008
AAT Asn	GTG Val	ATG Met	AGC Ser 340	Lys	GTT Val	CAA Gln	GGA Gly	GAA Glu 345	Val	AGA Arg	AAT Asn	CTA Leu	TTC Phe 350	GGT Gly	GAC Asp	1056
AAA Lys	GAT Asp	TTC Phe 355	Ile	AAC Asn	GAA Glu	GAT Asp	GAT Asp 360	Val	GAA Glu	AGC Ser	CTT Leu	CCT Pro 365	TAT Tyr	CTC Leu	AAA Lys	1104
GCA	GTG	GTG	AAG	GAG	ACA	TTA	AGA	TTA	TTC	CCA	CCT	TCA	CCA	CTA	CTT	1152

1657

									-2.6							
Ala	Val 370	Val	Lys	Glu	Thr	Leu 375	Arg	Leu	Phe	Pro	Pro 380	Ser	Pro	Leu	Leu	
TTG Leu 385	CCA Pro	AGG Arg	GTA Val	ACA Thr	ATG Met 390	GAA Glu	ACA Thr	TGC Cys	AAC Asn	ATA Ile 325	GAA Glu	GGG Gly	TAC Tyr	GAA Glu	ATT Ile 400	1200
CF.A Gln	GCC Ala	AAA Lys	ACT Thr	ATA Ile 405	GTG Val	CAT His	GTT Val	AAT Asn	GCA Ala 410	TGG Trp	GCC Ala	ATA Ile	GCA Ala	AGG Arg 415	GAC Asp	1248
CCT Pro	GAG Glu	AAT Asn	TGG Trp 420	GAA Glu	GAG Glu	CCT Pro	GAG Glu	AAA Lys 425	TTT Phe	TTC Phe	CCC Pro	GAA Glu	AGG Arg 430	TTC Phe	CTT Leu	1296
GAG Glu	AGT Ser	TCG Ser 435	ATG Met	GAG Glu	TTA Leu	AAG Lys	GGG Gly 440	AAT Asn	GAT Asp	GAG Glu	TTT Phe	AAG Lys 445	GTG Val	ATC Ile	CCG Pro	1344
TTT Phe	GGT Gly 450	TCT Ser	GGA Gly	AGG Arg	AGA Arg	ATG Met 455	TGT Cys	CCT Pro	GCG Ala	AAG Lys	CAC His 460	ATG Met	GGA Gly	ATT Ile	ATG Met	1392
AAT Asn 465	GTT Val	GAG Glu	CTT Leu	TCT Ser	CTT Leu 470	GCT Ala	AAT Asn	CTC	ATT	CAC His 475	ACG Thr	TTT Phe	GAT Asp	TGG Trp	GAA Glu 480	1440
GTG Val	GCT Ala	AAA Lys	GGG Gly	TTC Phe 485	GAC Asp	AAG Lys	GAA Glu	GAA Glu	ATG Met 490	TTG Leu	GAC Asp	ACG Thr	CAA Gln	ATG Met 495	AAA Lys	1488
CCA Pro	GGA Gly	ATA Ile	ACG Thr 500	ATG Met	CAC His	AAG Lys	AAA Lys	AGT Ser 505	GAT Asp	CTT Leu	TAC Tyr	CTA Leu	GTG Val 510	GCA Ala	AAG Lys	1536
		ACA Thr 515		TAG	CACA	CGT '	rgg t	ACAT	TC A	CTAT.	AACA	C AC.	AAGA	AAGT		1588
TGA'	TAAT	GAC '	TTGT	GTAT	GC A	ACTA	TGCT	C TA	TGCA	CTAT	GCA	CTAT	GTT '	TATT	GACCAT	1648

(2) INFORMATION FOR SEQ ID NO:12:

TAATTACTG

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 516 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Leu Val Leu Leu Ser Leu Leu Ser Ile Val Ile Ser Ile Val Leu Phe 1 5 10 15

Ile Thr His Thr His Lys Arg Asn Asn Thr Pro Arg Gly Pro Pro Gly 20 25 30

										•					
Pro	Pro	Pro .35	Leu	Pro	Leu	Ile	Gly 40	Asn	Leu	His	Gln	Leu 45	His	Asn	Ser
Ser	Pro 50	His	Leu	Cys	Leu	Trp 55	Gln	Leu	Ala	Lys	Leu 60	His	Gly	Pro	Leu
Met 65	Ser	Phe	Arg	Leu	Gly 70	Ala	Val	Gln	Thr	Val 75	Val	Val	Ser	Ser	Ala 80
Arg	Ile	Ala	Glu	Gln 85	Ile	Leu	ŗ'ns	Thr	His 90	Asp	Leu	Asn	Phe	Ala 95	Ser
Arg	Pro	Leu	Phe 100	Val	Gly	Pro	Arg	Lys 105	Leu	Ser	Tyr	Asp	Gly 110	Leu	Asp
Met	Gly	Phe 115	Ala	Pro	Tyr	Gly	Pro 120	Tyr	Trp	Arg	Glu	Met 125	Lys	Lys	Leu
Cys	Ile 130	Val	His	Leu	Phe	Ser 135	Ala	Gln	Arg	Val	Arg 140	Ser	Phe	Arg	Pro
Ile 145	Arg	Glu	Asn	Glu	Val 150	Ala	Lys	Met	Val	Arg 155	Lys	Leu	Ser	Glu	His 160
Glu	Ala	Ser	Gly	Thr 165	Val	Val	Asn	Leu	Thr 170	Glu	Thr	Leu	Met	Ser 175	Phe
Thr	Asn	Ser	Leu 180	Ile	Cys	Arg	Ile	Ala 185	Leu	Gly	Lys	Ser	Tyr 190	Gly	Cys
Glu	Tyr	Glu 195	Glu	Val	Val	Val	Asp 200	Glu	Val	Leu	Gly	Asn 205	Arg	Arg	Ser
Arg	Leu 210		Val	Leu	Leu	Asn 215		Ala	Gln	Ala	Leu 220	Leu	Ser	Glu	Phe
Phe 225	Phe	Ser	Asp	Tyr	Phe 230	Pro	Pro	Ile	Gly	Lys 235	Trp	Val	Asp	Arg	Val 240
Thr	Gly	Ile	Leu	Ser 245		Leu	. Asp	Lys	Thr 250	Phe	Lys	Glu	Leu	Asp 255	Ala
Cys	Tyr	Glu	Arg 260		Ser	Tyr	Asp	His 265	Met	. Asp	Ser	Ala	Lys 270	Ser	Gly
Lys	Lys	Asp 275	Asn	Asp	Asn	Lys	Glu 280		Lys	Asp	Ile	1le 285	Asp	Ile	Leu
Leu	Gln 290		Leu	Asp	Asp	Arg 295		Phe	Thr	Phe	a Asp 300	Leu	Thr	Leu	Asp
His 305		Lys	Ala	Val	Leu 310		: Asn	Ile	Phe	11e 315	Ala	Gly	Thr	Asp	Pro 320
Ser	Ser	Ala	Thr	1le 325		Trp) Ala	. Met	330		. Leu	Lev	Lys	335	Pro
Asn	Val	. Met	Ser		. Val	Glr	ı Gly	Glu 345		Arg	J Asr	Let	Phe 350	Gly	Asp

-28-

							_		0 1	0	T	Dwo	سره <i>د</i> يل	Len	Lvs	
		355					360					Pro 365				
Ala	Val 370	Val	Lys	Glu	Thr	Leu 375	Arg	Leu	Phe	Pro	Pro 380	Ser	Pro	Leu	Leu	
Leu 385	Pro	Arg	Val	Thr	Met 390	Glu	Thr	Cys	Asn	Ile 395	Glu	Gly	Tyr	Glu	Ile 400	
Gln	Ala	Lys	Thr	Ile 405	Val	His	Val	Asn	Ala 410	Trp	Ala	Ile	Ala	Arg 415	Asp	
Pro	Glu	Asn	Trp 420	Glu	Glu	Pro	Glu	Lys 425	Phe	Phe	Pro	Glu	Arg 430	Phe	Leu	
Glu	Ser	Ser 435	Met	Glu	Leu	Lys	Gly 440	Asn	Asp	Glu	Phe	Lys 445	Val	Ile	Pro	
Phe	Gly 450	Ser	Gly	Arg	Arg	Met 455	Cys	Pro	Ala	Lys	His 460	Met	Gly	Ile	Met	
Asn 465	.Val	Glu	Leu	Ser	Leu 470	Ala	Asn	Leu	Ile	His 475	Thr	Phe	Asp	Trp	Glu 480	
Val	Ala	Lys	Gly	Phe 485	Asp	Lys	Glu	Glu	Met 490	Leu	Asp	Thr	Gln	Met 495	Lys	
Pro	Gly	Ile	Thr 500	Met	His	Lys	Lys	Ser 505	Asp	Leu	Tyr	Leu	Val 510	Ala	Lys	
Lys	Pro	Thr 515														
(2)	INF	ORMA	TION	FOR	SEQ	ID	NO:1	3:								
	(i	(A) L B) T C) S	ENGT YPE : TRAN	HARA H: 1 nuc DEDN	824 leic ESS:	base aci sin	pai d	rs							
	(ii) MO	LECU	LE T	YPE:	cDN	Ά									
	(ix	:) FE (A) N	AME/	KEY :	CDS	.161	.6								
	(xi	.) SE	QUEN	ICE I	ESCR	IPTI	ON:	SEQ	ID N	10:13	:					
GG	AAAA:	TAG	CCTC	ACA!	AAA G	CAAF	GAT(CA AA	ACAA.	ACCAI	A GGI	ACGAG	AAC	ACG	ATG Met 1	56
TT(G CT u Lev	r GAZ ı Glu	ı Leı	GCA 1 Ala	A CTT a Lev	GG7	TTI Lev	A TTO Lev	ı Vaj	r TTC	G GC:	r CTC	TTT Phe	. Let	CAC His	104

-29-

									-29								
TTG Leu	CGT Arg	CCC Pro	ACA Thr	CCC Pro	ACT Thr	GCA Ala	AAA Lys 25	TCA Ser	AAA Lys	GCA Ala	CTT Leu	CGC Arg 30	CAT His	CTC Leu	CCA Pro		152
AAC Asn	CCA Pro 35	CCA Pro	AGC Ser	CCA Pro	AAG Lys	CCT Pro 40	CGT Arg	CTT Leu	CCC Pro	TTC Phe	ATA Ile 45	GGA Gly	CAC His	CTT Leu	CAT His		200
CTC Leu 50	TTA Leu	AAA Lys	GAC Asp	AAA Lys	CTT Leu 55	CTC Leu	CAC His	TAC Tyr	GCA Ala	CTC Leu 60	ATC Ile	GAC Asp	CTC Leu	TCC Ser	AAA Lys 65		248
AAA Lys	CAT His	GGT Gly	CCC Pro	TTA Leu 70	TTC Phe	TCT Ser	CTC Leu	TAC Tyr	TTT Phe 75	GGC Gly	TCC Ser	ATG Met	CCA Pro	ACC Thr 80	GTT Val	•	296
GTT Val	GCC Ala	TCC Ser	ACA Thr 85	CCA Pro	GAA Glu	TTG Leu	TTC Phe	AAG Lys 90	CTC Leu	TTC Phe	CTC Leu	CAA	ACG Thr 95	CAC His	GAG Glu		344
GCA Ala	ACT Thr	TCC Ser 100	TTC Phe	AAC Asn	ACA Thr	AGG Arg	TTC Phe 105	CAA Gln	ACC Thr	TCA Ser	GCC Ala	ATA Ile 110	AGA Arg	CGC Arg	CTC Leu		392
ACC Thr	TAT Tyr 115	GAT Asp	AGC Ser	TCA Ser	GTG Val	GCC Ala 120	ATG Met	GTT Val	CCC	TTC Phe	GGA Gly 125	Pro	TAC Tyr	TGG Trp	AAG Lys		440
TTC Phe 130	Val	AGG Arg	AAG Lys	CTC Leu	ATC Ile 135	ATG Met	AAC Asn	GAC Asp	CTT Leu	CCC Pro 140	Asn	GCC Ala	ACC Thr	ACT Thr	GTA Val 145	-	488
AAC Asn	AAG Lys	TTG Leu	AGG Arg	CCT Pro 150	Leu	AGG Arg	ACC Thr	CAA Gln	CAG Gln 155	Thr	CGC	: AAG Lys	TTC Phe	CTT Leu 160	AGG Arg		536
GTT Val	ATG Met	GCC Ala	CAA Gln 165	Gly	GCA Ala	GAG Glu	GCA Ala	CAG Gln 170	. Lys	CCC Pro	CTT Lev	GAC Asp	TTG Leu 175	III	GAG Glu		584
GAG Glu	CTT Leu	CTG Lev 180	. Lys	TGG	ACC	AAC Asn	AGC Ser 185	Thr	: ATC	TCC Ser	ATG Met	ATG Met	. Met	CTC Lev	GGC Gly		632
GAG Glu	GCT Ala 195	Gli	G GAG	ATC	: AGA : Arg	GAC Asp 200	Ile	GCT Ala	CGC Arg	GAG Glu	GTT Val 209	r rer	AAG Lys	ATC	TTT Phe		680
GG(Gl _y 21(/ Glu	TAC	AGC Ser	CTC	ACT Thr 215	Asp	TTC Phe	ATC	TGC Trp	CCA Pro 220) Let	J AAC	CAT His	CTC Lev	AAG Lys 225		728
GT: Val	r GGA L Gly	A AAC	TAT	GAC Glu 230	ı Lys	AGG Arg	ATO	GAC Asp	GAC Asi 23	o Ile	C TTO	J AAC u Asi	AAC Lys	TT0 Phe 240	GAC Asp		776
CC'	r GT(o Val	C GT l Vai	r GA 1 Glu 24!	ı Arç	G GTO	C ATO	AA(3 AA0 5 Lys 250	s Arg	C CG	r GAG	G ATO	GT(e Val 259	LAF	G AGG g Arg		824

									-30)						
AGA Arg	AAG Lys	AAC Asn 260	gga gly	GAG Glu	GTT Val	GTT Val	GAG Glu 265	GGT Gly	GAG Glu	GTC Val	AGC Ser	GGG Gly 270	GTT Val	TTC Phe	CTT Leu	872
GAC Asp	ACT Thr 275	TTC Leu	CTT Lau	GAA Glu	TTC Phe	GCT Ala 280	GAG Glu	GAT Asp	GAG Glu	ACC Thr	ATG Met 285	GAG Glu	ATC Ile	AAA Lys	ATC Ile	920
ACC Thr 290	Lys	GAC Asp	CAC His	ATC Ile	GAG Glu 295	GGT Gly	CTT	GTT Val	GTC Val	GAC Asp 300	TTT Phe	TTC Phe	TCG Ser	GCA Ala	GGA Gly 305	968
ACA Thr	GAC Asp	TCC Ser	ACA Thr	GCG Ala 310	GTG Val	GCA Ala	ACA Thr	GAG Glu	TGG Trp 315	GCA Ala	TTG Leu	GCA Ala	GAA Glu	CTC Leu 320	ATC Ile	1016
AAC Asn	AAT Asn	CCT Pro	AAG Lys 325	GTG Val	TTG Leu	GAA Glu	AAG Lys	GCT Ala 330	CGT Arg	GAG Glu	GAG Glu	GTC Val	TAC Tyr 335	AGT Ser	GTT Val	1064
GTG Val	GGA Gly	AAG Lys 340	GAC Asp	AGA Arg	CTT Leu	GTG Val	GAC Asp 345	GAA Glu	GTT Val	GAC Asp	ACT Thr	CAA Gln 350	AAC Asn	CTT Leu	CCT Pro	1112
TAC Tyr	ATT Ile 355	AGA Arg	GCA Ala	ATC Ile	GTG Val	AAG Lys 360	GAG Glu	ACA Thr	TTC Phe	CGC Arg	ATG Met 365	CAC His	CCG Pro	CCA Pro	CTC Leu	1160
CCA Pro 370	GTG Val	GTC Val	AAA Lys	AGA Arg	AAG Lys 375	TGC Cys	ACA Thr	GAA Glu	GAG Glu	TGT Cys 380	GIU	ATT	AAT Asn	GGA Gly	TAT Tyr 385	1208
GTG Val	ATC Ile	CCA Pro	GAG Glu	GGA Gly 390	Ala	TTG Leu	ATT Ile	CTC Leu	TTC Phe 395	AAT Asn	GTA Val	TGG Trp	CAA Gln	GTA Val 400	GGA Gly	1256
AGA Arg	GAC Asp	CCC Pro	AAA Lys 405	Tyr	TGG Trp	GAC Asp	AGA Arg	CCA Pro 410	Ser	GAG Glu	TTC Phe	CGT Arg	CCT Pro 415	GIU	AGG Arg	1304
TTC Phe	CTA Leu	GAG Glu 420	Thr	GGG Gly	GCT Ala	GAA Glu	GGG Gly 425	Glu	GCA Ala	. GGG Gly	CCT	CTT Leu 430	ASD	CTT Leu	AGG Arg	1352
GGA Gly	CAA Gln 435	His	TTT Phe	CAA Gln	CTT Leu	CTC Leu 440	Pro	TTT Phe	GGG Gly	TCT Ser	GGG Gly 445	Arg	AGA Arg	ATG Met	TGC Cys	1400
CCT Pro 450	Gly	GTC Val	: AAT . Asn	CTG Leu	GCT Ala 455	Thr	TCG Ser	GGA Gly	ATG Met	GCA Ala 460	i Thr	CTI Leu	CTI Leu	GCA Ala	TCT Ser 465	1448
CTT Leu	ATI	CAG Gln	TGC	TTC: Phe	Asp	TTG	CAA Glr	A GTG 1 Val	CTG Lev 475	r GI?	CCA Pro	A CAA o Glr	GGA Gly	CAG Gln 480	ATA Ile	1496
TTG Leu	AAG Lys	GGI Gly	GGT Gly 485	/ Asp	GCC Ala	AAA Lys	GTT Val	AGC Ser 490	: Met	GAA Glu	A GAC	AGA Arg	495	r GTA	CTC Leu	1544

	-31-	
ACT GTT CCA Thr Val Pro	A AGG GCA CAT AGT CTT GTC TGT GTT CCA CTT GCH HOS ATO D Arg Ala His Ser Leu Val Cys Val Pro Leu Ala Arg Ile	L592
GGC GTT GCA Gly Val Ala 515	TCT AAA CTC CTT TCT TAATTAAGAT CATCATCATA TATAATATTT 1 a Ser Lys Leu Leu Ser 520	1646
ACTTITTGTG	TGTTGATAAT CATCATTICA ATAAGGTCTC GTTCATCTAC TTTTTATGAA	1706
GTATATAAGC	CCTTCCATGC ACATTGTATC ATCTCCCATT TGTCTTCGTT TGCTACCTAA	1766
GGCAATCTTT	TTTTTTTTAG AATCACATCA TCCTACTATA AACTATCAAT CCTTATAT	1824
(2) INFORM	ATION FOR SEQ ID NO:14:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 521 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: protein	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

	(×	i) S	EQUE	NCE	DESC	.RIPI	LON	JEC	, 10	5 110.14.						
Met 1	Leu	Leu	Glu	Leu 5	Ala	Leu	Gly	Leu	Leu 10	Val	Leu	Ala	Leu	Phe 15	Leu	
His	Leu	Arg	Pro 20	Thr	Pro	Thr	Ala	Lys 25	Ser	Lys	Ala	Leu	Arg 30	His	Leu	
Pro	Asn	Pro 35	Pro	Ser	Pro	Lys	Pro 40	Arg	Leu	Pro	Phe	Ile 45	Gly	His	Leu	
His	Leu 50	Leu	Lys	Asp	Lys	Leu 55	Leu	His	Tyr	Ala	Leu 60	Ile	Asp	Leu	Ser	
Lys 65	Lys	His	Gly	Pro	Leu 70	Phe	Ser	Leu	Tyr	Phe 75	Gly	Ser	Met	Pro	Thr 80	
Val	Val	Ala	Ser	Thr 85	Pro	Glu	Leu	Phe	Lys 90	Leu	Phe	Leu	Gln	Thr 95	His	
Glu	Ala	Thr	Ser 100	Phe	Asn	Thr	Arg	Phe 105	Gln	Thr	Ser	Ala	Ile 110	Arg	Arg	
Leu	Thr	Tyr 115	Asp	Ser	Ser	Val	Ala 120	Met	Val	Pro	Phe	Gly 125	Pro	Tyr	Trp	
Lys	Phe 130	Val	Arg	Lys	Leu	Ile 135	Met	Asn	Asp	Leu	Pro 140	Asn	Ala	Thr	Thr	
Val 145		Lys	Leu	Arg	Pro 150	Leu	Arg	Thr	Gln	Gln 155	Thr	Arg	Lys	Phe	Leu 160	
Arg	Val	Met	Ala	Gln 165		Ala	Glu	Ala	Gln 170	Lys	Pro	Leu	Asp	Leu 175	Thr	

Glu Glu Leu Leu Lys Trp Thr Asn Ser Thr Ile Ser Met Met Leu

-32-

									٠.	_					
			180					185					190		
Gly	Glu	Ala 195	Glu	Glu	Ile	Arg	Asp 200	Ile	Ala	Arg	Glu	Val 205	Leu	Lys	Ile
Phe	Gly 210	Glu	Tyr	Ser	Leu	Thr 215	Asp	Phe	Ile	Trp	Pro 220	Leu	Lys	His	Leu
Lys 225	Val	Gl.y	Lys	Tyr	Glu 230	Lys	Arg	Ile	Asp	Asp 235	Ile	Leu	Asn	Lys	Phe 240
Asp	Pro	Val	Val	Glu 245	Arg	Val	Ile	Lys	Lys 250	Arg	Arg	Glu	Ile	Val 255	Arg
Arg	Arg	Lys	Asn 260	Gly	Glu	Val	Val	Glu 265	Gly	Glu	Val	Ser	Gly 270	Val	Phe
Leu	Asp	Thr 275	Leu	Leu	Glu	Phe	Ala 280	Glu	Asp	Glu	Thr	Met 285	Glu	Ile	Lys
Ile	Thr 290	Lys	Asp	His	Ile	Glu 295	Gly	Leu	Val	Val	Asp 300	Phe	Phe	Ser	Ala
Gly 305		Asp	Ser	Thr	Ala 310	Val	Ala	Thr	Glu	Trp 315	Ala	Leu	Ala	Glu	Leu 320
Ile	Asn	Asn	Pro	Lys 325	Val	Leu	Glu	Lys	Ala 330	Arg	Glu	Glu	Val	Tyr 335	Ser
Val	Val	Gly	Lys 340		Arg	Leu	Val	Asp 345	Glu	Val	Asp	Thr	Gln 350	Asn	Leu
		355					360					365			
Leu	Pro 370		Val	Lys	Arg	Lys 375	Cys	Thr	Glu	Glu	380	Glu	Ile	Asn	Gly
385			Pro		390					395					400
Gly	Arg	Asp	Pro	Lys 405	Tyr	Trp	Asp	Arg	Pro 410	Ser	Glu	Phe	Arg	Pro 415	Glu
			420					425	i				430		Leu
Arg	g Gly	Gln 435		Phe	Gln	Leu	Leu 440	Pro	Phe	: Gly	.Ser	Gly 445	Arg	Arg	Met
Cys	9 Pro		v Val	. Asn	Leu	Ala 455	Thr	Ser	Gly	Met	Ala 460	Thr	Leu	Leu	Ala
Ser 465		ı Ile	e Gln	Cys	Phe 470		Leu	glr	ı Val	. Lev 475	Gly	Pro	Gln	Gly	Gln 480
Ile	e Lev	ı Lys	s Gly	/ Gly 485		Ala	Lys	val	490	Met	: Glu	. Glu	a Arg	Ala 495	Gly
Lev	ı Thi	val	L Pro	Arg	Ala	His	s Ser	Le	ı Val	L Cys	val	Pro	Leu	Ala	Arg

-33-

500 505 510

Ile Gly Val Ala Ser Lys Leu Leu Ser 515 520

- (2) INFORMATION FOR SEQ ID NO:15:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1831 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: CDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 20..1747
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CAACACTCGC AGTACCGCC ATG AGT Met Ser	Val A	Asp Thr	Ser	Ser	Thr	Leu	Ser 10	Thr	52
--------------------------------------	-------	---------	-----	-----	-----	-----	-----------	-----	----

- GTC ACC GAT GCC AAT CTT CAC TCC AGA TTT CAT TCT CGT CTT GTT CCA 10

 Val Thr Asp Ala Asn Leu His Ser Arg Phe His Ser Arg Leu Val Pro

 15 20 25
- TTC ACT CAT CAT TTC TCA CTT TCT CAA CCC AAA CGG ATT TCT TCA ATC

 Phe Thr His His Phe Ser Leu Ser Gln Pro Lys Arg Ile Ser Ser Ile

 30 35 40
- AGA TGC CAA TCA ATT AAT ACC GAT AAG AAG AAA TCA AGT AGA AAT CTG

 Arg Cys Gln Ser Ile Asn Thr Asp Lys Lys Lys Ser Ser Arg Asn Leu

 50

 55
- CTG GGC AAT GCA AGT AAC CTC CTC ACG GAC TTA TTA AGT GGT GGA AGT
 Leu Gly Asn Ala Ser Asn Leu Leu Thr Asp Leu Leu Ser Gly Gly Ser
 60 70 75
- ATA GGG TCT ATG CCC ATA GCT GAA GGT GCA GTC TCA GAT CTG CTT GGT

 11e Gly Ser Met Pro Ile Ala Glu Gly Ala Val Ser Asp Leu Leu Gly

 80

 85

 90
- CGA CCT CTC TTT TTC TCA CTG TAT GAT TGG TTC TTG GAG CAT GGT GCG

 Arg Pro Leu Phe Phe Ser Leu Tyr Asp Trp Phe Leu Glu His Gly Ala

 95 100 105
- GTG TAT AAA CTT GCC TTT GGA CCA AAA GCA TTT GTT GTT GTA TCA GAT
 Val Tyr Lys Leu Ala Phe Gly Pro Lys Ala Phe Val Val Val Ser Asp
 110 115 120
- CCC ATA GTT GCT AGA CAT ATT CTG CGA GAA AAT GCA TTT TCT TAT GAC

 Pro Ile Val Ala Arg His Ile Leu Arg Glu Asn Ala Phe Ser Tyr Asp

 125

 130

 135
- AAG GGA GTA CTT GCT GAT ATC CTT GAA CCA ATA ATG GGC AAA GGA CTC 484

-34-

									-54-	_						
Lys 140					145					150						
ATA Ile	CCA Pro	GCA Ala	GAC Asp	CTT Leu 160	GAT Asp	ACT Thr	TGG Trp	AAG Lys	CAA Gln 165	AGG Arg	AGA Arg	AGA Arg	GTC Val	ATT Ile 170	GCT Ala	532
CCG Pro	GCT Ala	TTC Phe	CAT His 175	AAC Asn	TCA Ser	TAC Tyr	TTG Leu	GAA Glu 180	GCT Ala	ATG Met	GTT Val	AAA Lys	ATA Ile 185	TTC Phe	ACA Thr	580
ACT Thr	TGT Cys	TCA Ser 190	GAA Glu	AGA Arg	ACA Thr	ATA Ile	TTG Leu 195	AAG Lys	TTT Phe	AAT Asn	AAG Lys	CTT Leu 200	CTT Leu	GAA Glu	GGA Gly	628
GAG Glu	GGT Gly 205	TAT Tyr	GAT Asp	GGA Gly	CCT Pro	GAC Asp 210	TCA Ser	ATT Ile	GAA Gļu	TTG Leu	GAT Asp 215	CTT Leu	GAG Glu	GCA Ala	GAG Glu	676
TTT Phe 220	TCT Ser	AGT Ser	TTG Leu	GCT Ala	CTT Leu 225	GAT Asp	ATT Ile	ATT Ile	GGG Gly	CTT Leu 230	GGT Gly	GTG Val	TTC Phe	AAC Asn	TAT Tyr 235	724
GAC Asp	TTT Phe	GGT Gly	TCT Ser	GTC Val 240	Thr	AAA Lys	GAA Glu	TCT	CCA Pro 245	GTT Val	ATT Ile	AAG Lys	GCA Ala	GTC Val 250	TAT Tyr	772
GGC Gly	ACT Thr	CTT Leu	TTT Phe 255	Glu	GCT Ala	GAA Glu	CAC His	AGA Arg 260	TCC Ser	ACT Thr	TTC Phe	TAC Tyr	ATT Ile 265	CCA Pro	TAT	820
TGG Trp	AAA Lys	ATT Ile	Pro	TTG Leu	GCA Ala	AGG Arg	TGG Trp 275	ile	GTC Val	CCA Pro	AGG Arg	CAA Gln 280	Arg	AAG Lys	TTT Phe	868
CAG Gln	GAT Asp 285	Asp	CTA Leu	. AAG Lys	GTC Val	Ile 290	Asn	ACT Thr	TGT	CTT Leu	GAT Asp 295	, Gry	CTT Leu	ATC Ile	AGA Arg	916
AAT Asn 300	Ala	AAA Lys	A GAG	AGC Ser	AGA Arg	Gln	GAA Glu	ACA Thr	GAT Asp	GTT Val	. 610	AAA Lys	TTG Leu	CAG Gln	CAG Gln 315	964
AGG Arg	GAT As <u>r</u>	TAC Ty:	TTA	A AAT A AST 320	ı Lev	AAG Lys	GAT Asp	GCA Ala	AGT Ser 325	Lev	CTC Lev	G CGT	TTC Phe	CTC Lev 330	GTT Val	1012
GAT Asr	ATC Met	CG(G GGF G Gly 335	/ Ala	r GAT a Asp	GTT Val	GA:	GAT ASE 340	Arç	CAC Glr	TT(AGO Arg	GAT J ASP 345		TTA Leu	1060
AT(Met	ACA Thi	A ATO	t Lev	r AT	r GC0 e Ala	GGT a Gly	CA' His	s GIV	A ACA	A ACC	G GC	r GCA a Ala 360	a val	CT?	r ACT 1 Thr	110
TG(Tr]	G GCI o Ala 36!	a Va	T TTO	C CTO	C CT	A GCT a Ala 370	a GL	A AA? n Asi	r CC	r AGG	C AA Ly 37	s Me	E Lys	AAG ELY	G GCT S Ala	115

-35-

									-35	_							•
CAA Gln 380	GCA Ala	GAG Glu	GTA Val	GAT Asp	TTG Leu 385	GTG Val	CTG Leu	GGT Gly	ACG Thr	GGG Gly 390	AGG Arg	CCA Pro	ACT Thr	TTT Phe	GAA Glu 395	12	04
TCA Ser	CTT Leu	AAG Lys	GAA Glu	TTG Leu 400	CAG Gln	TAC Tyr	ATT Ile	AGA Arg	TTG Leu 405	ATT Ile	GTT Val	GTG Val	GAG Glu	GCT Ala 410	CTT Leu	12	52
CGT Arg	TTA Leu	TAC Tyr	CCC Pro 415	CAA Gln	CCA Pro	CCT Pro	TTG Leu	CTG Leu 420	ATT Ile	AGA Arg	CGT Arg	TCA Ser	CTC Leu 425	AAA Lys	TCT Ser	13	00
GAT Asp	GTT Val	TTA Leu 430	CCA Pro	GGT Gly	GGG Gly	CAC His	AAA Lys 435	GGT Gly	GAA Glu	AAA Lys	GAT Asp	GGT Gly 440	TA.T Tyr	GCA Ala	ATT Ile	13	48
CCT Pro	GCT Ala 445	GGG Gly	ACT Thr	GAT Asp	GTC Val	TTC Phe 450	ATT Ile	TCT Ser	GTA Val	TAT Tyr	AAT Asn 455	CTC Leu	CAT His	AGA Arg	TCT Ser	13	196
CCA Pro 460	TAT Tyr	TTT Phe	TGG Trp	GAC Asp	CGC Arg 465	CCT Pro	GAT	GAC Asp	TTC Phe	GAA Glu 470	CCA Pro	GAG Glu	AGA Arg	TTT Phe	CTT Leu 475	14	144
GTG Val	CAA Gln	AAC Asn	AAG Lys	AAT Asn 480	GAA Glu	GAA Glu	ATT	GAA Glu	GGA Gly 485	Lrp	GCT Ala	GGT Gly	CTT Leu	GAT Asp 490	CCA Pro	14	192
TCT Ser	CGA Arg	AGT Ser	CCC Pro 495	Gly	GCC Ala	TTG Leu	TAT	CCG Pro 500	Asn	GAG Glu	GTT Val	ATA Ile	TCG Ser 505	ASP	TTT Phe	15	540
GCA Ala	TTC Phe	TTA Leu 510	Pro	TTT Phe	GGT Gly	GGC	GGA Gly 515	Pro	CGA Arg	AAA Lys	TGT Cys	GTT Val 520	GGG Gly	GAC Asp	CAA Gln	1!	588
TTT Phe	GCT Ala 525	. Leu	ATG Met	GAG Glu	TCC Ser	ACT Thr 530	Val	GCG Ala	TTC Lev	ACT Thr	ATG Met	Leu	CTC Leu	CAG Gln	AAT Asn	1	636
TTT Phe 540	Asp	GTG Val	GAA Glu	CTA Leu	AAA Lys 545	Gly	ACC Thr	CCT Pro	GAZ Glu	TCG Ser 550	· Vai	GAA Glu	. CTA . Leu	GTT Val	ACT Thr 555	1	684
GGG Gly	GCA Ala	A ACT	T ATT	CAT His	Thr	Lys	AAT : Asr	r GGA n Gly	A ATO	Tr	TGC Cys	AGA Arg	TTG Leu	AAG Lys 570	AAG Lys	1	732
			r TT/ 1 Lev 579	ı Arg		ACATA	ATGT	ACTO	GTGG(CA 1	TTTT	rctt <i>i</i>	OA TA	AGAA	TAAT	1	787
GTA	ATAT:	TATT	ATT	CTTTC	GAG A	ATA	TAT	GA A	raaa'	TTCC'	r aga	AC				1	831

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 576 amino acids

-36-

- (B) TYPE: amino acid
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Ser Val Asp Thr Ser Ser Thr Leu Ser Thr Val Thr Asp Ala Asn

Leu His Ser Arg Phe His Ser Arg Leu Val Pro Phe Thr His His Phe 20 25 30

Ser Leu Ser Gln Pro Lys Arg Ile Ser Ser Ile Arg Cys Gln Ser Ile 35 40 45

Asn Thr Asp Lys Lys Ser Ser Arg Asn Leu Leu Gly Asn Ala Ser 50 60

Asn Leu Leu Thr Asp Leu Leu Ser Gly Gly Ser Ile Gly Ser Met Pro 65 70 75 80

Ile Ala Glu Gly Ala Val Ser Asp Leu Leu Gly Arg Pro Leu Phe Phe 85 90 95

Ser Leu Tyr Asp Trp Phe Leu Glu His Gly Ala Val Tyr Lys Leu Ala 100 105 110

Phe Gly Pro Lys Ala Phe Val Val Val Ser Asp Pro Ile Val Ala Arg 115 120 125

His Ile Leu Arg Glu Asn Ala Phe Ser Tyr Asp Lys Gly Val Leu Ala 130 135 140

Asp Ile Leu Glu Pro Ile Met Gly Lys Gly Leu Ile Pro Ala Asp Leu 145 150 155 160

Asp Thr Trp Lys Gln Arg Arg Arg Val Ile Ala Pro Ala Phe His Asn 165 170 175

Ser Tyr Leu Glu Ala Met Val Lys Ile Phe Thr Thr Cys Ser Glu Arg 180 185 190

Thr Ile Leu Lys Phe Asn Lys Leu Leu Glu Gly Glu Gly Tyr Asp Gly 195 200 205

Pro Asp Ser Ile Glu Leu Asp Leu Glu Ala Glu Phe Ser Ser Leu Ala 210 215 220

Leu Asp Ile Ile Gly Leu Gly Val Phe Asn Tyr Asp Phe Gly Ser Val 225 230 235 240

Thr Lys Glu Ser Pro Val Ile Lys Ala Val Tyr Gly Thr Leu Phe Glu 245 250 255

Ala Glu His Arg Ser Thr Phe Tyr Ile Pro Tyr Trp Lys Ile Pro Leu 260 265 270

Ala Arg Trp Ile Val Pro Arg Gln Arg Lys Phe Gln Asp Asp Leu Lys 275 280 285

- Val Ile Asn Thr Cys Leu Asp Gly Leu Ile Arg Asn Ala Lys Glu Ser Arg Gln Glu Thr Asp Val Glu Lys Leu Gln Gln Arg Asr Tyr Leu Asn 315 Leu Lys Asp Ala Ser Leu Leu Arg Phe Leu Val Asp Met Arg Gly Ala Asp Val Asp Asp Arg Gln Leu Arg Asp Asp Leu Met Thr Met Leu Ile 345 Ala Gly His Glu Thr Thr Ala Ala Val Leu Thr Trp Ala Val Phe Léu 360 Leu Ala Gln Asn Pro Ser Lys Met Lys Lys Ala Gln Ala Glu Val Asp Leu Val Leu Gly Thr Gly Arg Pro Thr Phe Glu Ser Leu Lys Glu Leu 390 Gln Tyr Ile Arg Leu Ile Val Val Glu Ala Leu Arg Leu Tyr Pro Gln 405 Pro Pro Leu Leu Ile Arg Arg Ser Leu Lys Ser Asp Val Leu Pro Gly Gly His Lys Gly Glu Lys Asp Gly Tyr Ala Ile Pro Ala Gly Thr Asp Val Phe Ile Ser Val Tyr Asn Leu His Arg Ser Pro Tyr Phe Trp Asp 455 Arg Pro Asp Asp Phe Glu Pro Glu Arg Phe Leu Val Gln Asn Lys Asn 470 Glu Glu Ile Glu Gly Trp Ala Gly Leu Asp Pro Ser Arg Ser Pro Gly 485 Ala Leu Tyr Pro Asn Glu Val Ile Ser Asp Phe Ala Phe Leu Pro Phe Gly Gly Gly Pro Arg Lys Cys Val Gly Asp Gln Phe Ala Leu Met Glu 520 Ser Thr Val Ala Leu Thr Met Leu Leu Gln Asn Phe Asp Val Glu Leu 535 Lys Gly Thr Pro Glu Ser Val Glu Leu Val Thr Gly Ala Thr Ile His 550 545 Thr Lys Asn Gly Met Trp Cys Arg Leu Lys Lys Arg Ser Asn Leu Arg 570
 - (2) INFORMATION FOR SEQ ID NO:17:

565

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1704 base pairs

-38-

(B)	TYPE: nucleic acid
(C)	STRANDEDNESS: single
(D)	TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE: (A) NAME/KEY: CDS

(B) LOCATION: 38..1564

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:												
CAGGCTCCA	C AAAAC	ATCTC AT	CATTC	ACC CAA	.CAAA	ATG Met 1	GCG Ala	CTG Leu	CTT	CTG Leu 5	ATA Ile	55
ATT CCC A	TC TCA le Ser 10	CTG GTC Leu Val	ACC C'	TC TGG eu Trp 15	CTC	GGT ' Gly '	TAC Tyr	ACC Thr	CTA Leu 20	TAC Tyr	CAG Gln	103
CGA TTA A	GA TTC rg Phe 25	AAG CTC Lys Leu	Pro P	CG GGT ro Gly	CCA Pro	CGG Arg	CCC Pro	TGG Trp 35	CCG Pro	GTA Val	GTC Val	151
GGT AAC C Gly Asn L 40	TC TAC	GAC ATA Asp Ile	AAA C Lys P 45	CC GTC ro Val	CGC Arg	TTC Phe	CGG Arg 50	TGC Cys	TTC Phe	GCG Ala	GAG Glu	199
TGG GCG C Trp Ala G 55	AG TCT	TAC GGC Tyr Gly 60	CCC A Pro I	TA ATA le Ile	TCG Ser	GTT Val 65	TGG Trp	TTC Phe	GGT Gly	TCG Ser	ACC Thr 70	247
CTA AAC G Leu Asn V	TC ATC	GTT TCG Val Ser 75	AAC T Asn S	CG GAG er Glu	CTG Leu 80	GCG Ala	AAG Lys	GAG Glu	GTG Val	CTG Leu 85	AAG Lys	295
GAG CAC G	AT CAG Asp Gln 90	CTG CTG Leu Leu	GCG G Ala A	AC CGC Asp Arg 95	CAC His	CGG Arg	AGC Ser	CGG Arg	TCG Ser 100	GCG Ala	GCG Ala	343
AAG TTC A	AGC CGC Ser Arg 105	GAC GGG Asp Gly	Lys A	SAT CTA Asp Leu 110	ATT	TGG Trp	GCC Ala	GAT Asp 115	TAT Tyr	GGG Gly	CCG Pro	391
CAC TAC (His Tyr \ 120	GTG AAG Val Lys	GTG AGG Val Arg	AAG C Lys V 125	GTT TGC Val Cys	ACG Thr	CTC Leu	GAG Glu 130	CTT Leu	TTC Phe	TCG Ser	CCG Pro	439
AAG CGC (Lys Arg) 135	CTC GAG Leu Glu	GCC CTG Ala Leu 140	Arg I	CCC ATT Pro Ile	AGG Arg	GAG Glu 145	GAC Asp	GAG Glu	GTC Val	ACC Thr	TCC Ser 150	487
ATG GTT (GAC TCC Asp Ser	GTT TAC Val Tyr 155	AAT (CAC TGO	ACC Thr	261	ACT Thr	GAA Glu	AAT Asn	TTG Leu 165	GGG Gly	535
AAA GGA Lys Gly	ATA TTG Ile Leu	TTG AGG	AAG (CAC TTO	GGG Gly	GTT Val	GTG Val	GCA Ala	TTC Phe	AAC Asn	AAC Asn	583

-39-

-39-																
			170					175					180			
ATA I	Thr	Arg 185	Leu	Ala	Phe	GIÀ	Lys 190	Arg	Pile	Val	ASII	195	014	1		631
	Asp 200	Glu	Gln	Gly	Val	205	Pne	гĀг	ALG	116	219	UI u		1		679
AAG Lys 215	Leu	Gly	Ala	Ser	Leu 220	Ala	Met	Ala	GIU	225	116	FLO	115		230	727
Trp	Met	Phe	Pro	Leu 235	GAA Glu	Glu	GIĀ	Ala	240	AIA	пуз	***	017	245	3	775
Arg	Asp	Arg	Leu 250	Thr	AGA Arg	Ala	IIe	мес 255	Ala	GIU	піз	1111	260	,,,,	,	823
Lys	Lys	Ser 265	Gly	Gly	Ala	Lys	270	HIS	Pne	Val	ASP	275	Lea	200		871
Leu	Gln 280	Asp	Lys	Tyr	Asp	Leu 285	Ser	GIU	Asp	Ini	290)	. Gir	200	CTT	919
Trp 295	Asp	Met	. Ile	Thr	300	Gly	Met	Asp	Tur	305	HIC	1 110			GAG Glu 310	967
Trp	Ala	Met	Ala	315	Leu S	Ile	Arg	Asn	320)	y va.	L GI	1 011	325		1015
CAA Gln	GAC Glu	GAC Glv	G CTA 1 Let 330	ı Ası	AGG Arg	Val	тте	: GT	ישני	r GA <i>l</i> ı Glı	1 21	G GTO	340		GAA Glu	1063
GCA Ala	GA0	7 TTC 9 Pho 34	e Se	A AAT	r CTC	CCI	TAC Tyr 350	. Let	A CAI	A TGʻ n Cy:	r GT s Va	G AC 1 Th 35	L Ly	A GAA	A GCA 1 Ala	1111
ATG Met	AGG Arg	g Le	r CA	C CC	A CCA	A ACC 5 Thi 365	. Pro	A CTI	A ATO	G CT	C CC u Pr 37	0 111	C CG s Ar	T GCC g Ala	a Asn	1159
GCC Ala 375	a As	T GT n Va	C AA 1 Ly	A GT s Va	r GG 1 Gl; 38	y Gl	TAT	r GA	C AT p Il	T CC e Pr 38	Ory	A GG s Gl	G TC y Se	C AA' r As	r GTG n Val 390	1207
CAT His	r GT s Va	G AA 1 As	T GT n Va	G TG 1 Tr 39	p Al	G GT a Va	G GCO	C CG a Ar	C GA g As 40	b br	G GC	C GI .a Va	G TG	G AA p Ly 40	G GAT s Asp 5	1255
CC	A TT o Le	G GA	G TI u Ph	C CG	A CC	C GA o Gl	A AG u Ar	G TT g Ph	C CI	T GA	G GA	AG GA Lu As	T GI p Va	'A GA il As	C ATG p Met	1303

				~40~			
	410		415			420	
AAG GGC CAT Lys Gly His 425	GAC TTT Asp Phe	Arg Leu L	TT CCA eu Pro 30	TTC GGG Phe Gly	TCG GGT Ser Gly 435	CGA CGA Arg Arg	GTA 1351 Val
TGC CCG GGT Cys Pro Gly 440	GCC CAA Ala Gln	CTT GGT A Leu Cly I 445	TC AAC le Asn	TTG GCA Leu Ala	GCA TCC Ala Ser 450	ATG TTG Met Leu	GGC 1399 Gly
CAC CTC TTG His Leu Leu 455	His His	Phe Cys T 460	rp Thr	465	GIU GIY	Met bys	470
GAG GAA ATT Glu Glu Ile	GAC ATG Asp Met 475	GGA GAG A Gly Glu A	AT CCA Asn Pro	GGG CTA Gly Leu 480	GTC ACA Val Thr	TAC ATG Tyr Met 485	AGG 1495 Arg
ACT CCA ATA	CAA GCT Gln Ala 490	GTG GTT T Val Val S	CCT CCT Ser Pro 495	AGG CTC Arg Leu	CCC TCA Pro Ser	CAT TTA His Leu 500	TAC 1543 Tyr
AAA CGT GTG Lys Arg Val	Pro Ala	GAG ATC T	TAATCTT	CT TTTC	TTTCCC T	rggactaci	1594
CTTTGTTGCA	TTAAGAAA	AA TGCCTTG	STGG CAG	CTACTTTT	ATCTTTG	TGT TTATO	STAACT 1654
ACATATGAAA	TCACAATT	TA AGGAACI	raag gaa	AAAACTCA	TTGCGAG	GGT	1704
(2) INFORMA	SEQUENCE (A) LE	SEQ ID NO CHARACTER NGTH: 509 PE: amino	RISTICS amino a	: acids			
		POLOGY: li					
		TYPE: pro					
(xi)	SEQUENCE	DESCRIPT	ION: SE	Q ID NO:	18:		
Met Ala Lev 1	5			10		12	
Gly Tyr Th	Leu Tyr 20	Gln Arg I	Leu Arg 25	Phe Lys	Leu Pro	30	PIO
Arg Pro Try		Val Gly	Asn Leu 40	Tyr Asp	Ile Lys 45	Pro Val	Arg
Phe Arg Cy. 50	s Phe Ala	Glu Trp 2 55	Ala Gln	Ser Tyr	60 Gly Pro	lle Ile	Ser
Val Trp Ph 65	e Gly Ser	Thr Leu . 70	Asn Val	Ile Val	Ser Asr	ı Ser Glu	Leu 80
Ala Lys Gl	u Val Lev 85		His Asp	Gln Lev 90	Leu Ala	Asp Arg 95	His

									-41						
Arg	Ser	Arg	Ser 100	Ala	Ala	Lys	Phe	Ser 105	Arg	Asp	Gly	Lys	Asp 110	Leu	Ile
Trp	Ala	Asp 115	Tyr	Gly	Pro	His	Tyr 120	Val	Lys	Val	Arg	Lys 125	Val	Cys	Thr
Leu	Glu 130	Leu	Phe	Ser	Pro	Lys 135	Arg	Leu	Glu	Ala	Leu 140	Arg	Pro	Ile	Arg
Glu 145	Asp	Glu	Val	Thr	Ser 150	Met	Val	Asp	Ser	Val 155	Tyr	Asn	His	Cys	Thr 160
Ser	Thr	Glu	Asn	Leu 165	Gly	Lys	Gly	Ile	Leu 170	Leu	Arg	Lys	His	Leu 175	Gly
Val	Val	Ala	Phe 180	Asn	Asn	Ile	Thr	Arg 185	Leu	Ala	Phe	Gly	Lys 190	Arg	Phe
Val	Asn	Ser 195	Glu	Gly	Val		Asp 200	Glu	Gln	Gly	Val	Glu 205	Phe	Lys	Ala
Ile	Val 210	Glu	Asn	Gly	Leu	Lys 215	Leu	Gly	Ala	Ser	Leu 220	Ala	Met	Ala	Glu
His 225	Ile	Pro	Trp	Leu	Arg 230	Trp	Met	Phe	Pro	Leu 235	Glu	Glu	Gly	Ala	Phe 240
Ala	Lys	His	Gly	Ala 245	Arg	Arg	Asp	Arg	Leu 250	Thr	Arg	Ala	Ile	Met 255	Ala
Glu	His	Thr	Glu 260	Ala	Arg	Lys	Lys	Ser 265	Gly	Gly	Ala	Lys	Gln 270	His	Phe
Val	Asp	Ala 275	Leu	Leu	Thr	Leu	Gln 280	Asp	Lys	Tyr	Asp	Leu 285	Ser	Glu	Asp
Thr	Ile 290	Ile	Gly	Leu	Leu	Trp 295	Asp	Met	Ile	Thr	Ala 300	Gly	Met	Asp	Thr
Thr 305		Ile	Ser	Val	Glu 310		Ala	Met	Ala	Glu 315	Leu	Ile	Arg	Asn	Pro 320
Arg	Val	Gln	Gln	Lys 325	Val	Gln	Glu	Glu	Leu 330	Asp	Arg	Val	Ile	Gly 335	Leu
Glu	Arg	Val	Met 340		Glu	Ala	Asp	Phe 345	Ser	Asn	Leu	Pro	Tyr 350	Leu	Gln
Cys	Val	Thr 355	_	Glu	Ala	Met	Arg 360	Leu	His	Pro	Pro	Thr 365	Pro	Leu	Met
Leu	Pro 370		Arg	Ala	Asn	Ala 375	Asn	Val	Lys	Val	Gly 380	Gly	Tyr	Asp	Ile
Pro 385	-	Gly	Ser	Asn	Val 390	His	Val	Asn	Val	Trp 395		Val	Ala	Arg	Asp 400
Pro	Ala	Val	Trp	Lys 405		Pro	Leu	Glu	Phe	Arg	Pro	Glu	Arg	Phe 415	Leu

-42-

- Glu Glu Asp Val Asp Met Lys Gly His Asp Phe Arg Leu Leu Pro Phe 420 425 430
- Gly Ser Gly Arg Arg Val Cys Pro Gly Ala Gln Leu Gly Ile Asn Leu
 435
 440
 445
- Ala Ala Ser Met Leu Gly His Leu Leu His His Phe Cys Trp Thr Pro
- Pro Glu Gly Met Lys Pro Glu Glu Ile Asp Met Gly Glu Asn Pro Gly
 465 470 475 480
- Leu Val Thr Tyr Met Arg Thr Pro Ile Gln Ala Val Val Ser Pro Arg
 485 490 495
- Leu Pro Ser His Leu Tyr Lys Arg Val Pro Ala Glu Ile 500 505
- (2) INFORMATION FOR SEQ ID NO:19:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TGTCTAACTC CTTCCTTTTC

20

- (2) INFORMATION FOR SEQ ID NO:20:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
 - Phe Leu Pro Phe Gly Xaa Gly Xaa Arg Xaa Cys Xaa Gly
- (2) INFORMATION FOR SEQ ID NO:21:

-43-

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Phe Xaa Xaa Gly Xaa Xaa Cys Xaa Gly

- (2) INFORMATION FOR SEQ ID NO:22:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Xaa Cys Xaa Gly

- (2) INFORMATION FOR SEQ ID NO:23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Pro Glu Glu Phe Xaa Pro Glu Arg Phe 5

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/53 C12N C12N15/82 C12N9/02 C12N5/00 A01H5/00 A01H5/10 According to International Patent Classification (IPC) or to both national classification and IPC B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12N A01H Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Category 3 Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No "Induction of a novel SUZUKI G ET AL: Α 1-47 cytochrome P450 (CYP93 family) b methyl jasmonate in soybean suspension-cultured cells." FEBS LETTERS, (1996 MAR 25) 383 (1-2) 83-6. JOURNAL CODE: EUH. ISSN: 0014-5793., XP002046657 Nether lands see the whole document DATABASE WPI Α 1 - 47Section Ch, Week 9745 Derwent Publications Ltd., London, GB; Class C12, Page 10, AN 97-484100 XP002100401 & JP 09 224671 A (MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO), 2 September 1997 see abstract Further documents are listed in the continuation of box C. Χ Patent family members are listed in annex. Special categories of cited documents: "T" later document published after the international filing date or priority date and not in conflict with the application but "A" document defining the general state of the lart which is not considered to be of particular relevance cited to understand the principle or theory underlying the invention "E" earlier document but published on or after the international "X" document of particular relevance; the claimed invention filing date cannot be considered novel or cannot be considered to "L" document which may throw doubts on priority claim(s) or involve an inventive step when the document is taken alone hich is cited to establish the publication date of another "Y" document of particular relevance; the claimed invention citation or other special reason (as specified) cannot be considered to involve an inventive step when the document is combined with one or more other, such docu-"O" document referring to an oral disclosure, use, exhibition or ments, such combination being obvious to a person skilled other means "P" document published prior to the international filing date but later than the priority date claimed "&" document member of the same patent family Date of the actual completion of the international search Date of mailing of the international search report 20 April 1999 03/05/1999

Name and mailing address of the ISA

NL - 2280 HV Rijswijk

Fax: (+31-70) 340-3016

European Patent Office, P.B. 5818 Patentiaan 2

Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,

Authorized officer

Kania, T

1

INTERN ONAL SEARCH REPORT

.ational Application No PCT/US 98/20807

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT							
Category ,	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.					
A	WO 91 03561 A (DU PONT) 21 March 1991 cited in the application see the whole document	1-47					
A	FRANK M. ET AL.: "Cloning of wound-induced cytochrome P450 monooxygenases expressed in pea" PLANT PHYSIOLOGY, vol. 110, 1996, pages 1035-1046, XP002100394 see the whole document	1-47					
A	SHIOTA N. ET AL.: "Herbicide-resistant tobacco plants expressing the fused enzyme between rat cytochrome P4501A1 (CYP1A1) and yeast NADPH-cytochrome P450 oxidoreductase" PLANT PHYSIOLOGY, vol. 106, 1994, pages 17-23, XP002100395 cited in the application see the whole document	1-47					
Α	PIERREL M. ET AL.: "Catalytic properties of the plant cytochrome P450 CYP73 expressed in yeast" EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 224, no. 3, 1994, pages 835-844, XP002100396 cited in the application see the whole document	1-47					
A	KOCHS G. ET AL.: "Further characterization of cytochrome P450 in phytoalexin synthesis in soybean: cytochrome P450 cinnamate 4-hydroxylase and 3,9-dihydroxypterocarpan 6a-hydroxylase" ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 293, no. 1, 1992, pages 187-194, XP002100397 see the whole document	1-47					
P , X	SIMINSZKY B. ET AL.: "AC AF022157" EMBL DATABASE,8 January 1998, XP002100398 see the whole document	1-47					
Т	SCHOPFER C R ET AL: "Identification of elicitor-induced cytochrome P450s of soybean Glycine max L.) using differential display of mRNA." MOLECULAR AND GENERAL GENETICS, (1998 MAY) 258 (4) 315-22. JOURNAL CODE: NGP. ISSN: 0026-8925., XP002100399 GERMANY: Germany, Federal Republic of see the whole document	1-47					
	-/						

1



In Application No
PCT/US 98/20807

		PC1/US 98/	20007				
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT							
Category '	Citation of document, with indication where appropriate, of the relevant passages		Relevant to claim No.				
Т	SIMINSZKY B ET AL: "Expression of a soybean cytochrome P450 monopxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1999 FEB 16) 96 (4) 1750-5. JOURNAL CODE: PV3. ISSN: 0027-8424., XP002100400 United States see the whole document		1-47				
			7				
			•				
		:					
			/				



ť

INTERNATIONAL SEARCH REPORT

PCT/US 98/20807

Box I Observations where certain claims wire found unsearchable (Continuation of item 1 of first sheet)
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
Claims Nos.: Lecause they relate to subject matter not required to be searched by this Authority, namely:
2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
See additional sheet
As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest The additional search fees were accompanied by the applicant's protest.
No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (1)) (July 1998)

1:

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

- 1. Claims: 1-3,7-16 partially; 4-6,17-47 completely
 An isolated DNA molecule comprising a sequence consisting of SEQ ID NO:1,
 coding for an enzyme of SEQ ID NO:2, DNA sequences at least 90 % identical
 thereto and encoding a cytochrome P450 enzyme, and variants thereof. Encoded
 peptides, P450 enzymes, DNA constructs therewith, plant cells and transgenic
 plants comprising said constructs. A method of making a transgenic plant cell
 having an increased ability to metabolize phenylurea compounds compared to an
 untransformed cell, by transformation with said construct, and plants having
 increased resistance to phenylurea herbicides compared to wild-type plants of
 the same species, progeny and seed thereof. A crop comprising said plants. A
 method of using a phenylurea herbicide as a post-emergence herbicide,
 comprising planting said crop and applying a phenylurea herbicide thereto.
- 2. Claims: 1-3,7-16 partially
 An isolated DNA molecule comprising a sequence consisting of SEQ NO:3, coding
 for an enzyme of SEQ IS NO: 4, DNA sequences at least 90% identical thereto and
 encoding a cytochrome P450 enzyme, and variants thereof. Encoded peptides,
 P450, DNA constructs therewith, plant cells and transgenic plants comprising
 said constructs.
- 3. Claims: 1-3,7-16 partially idem for SED ID NOs: 5,6
 4. Claims: 1-3,7-16 partially idem for SED ID NOs: 7,8
 5. Claims: 1-3,7-16 partially idem for SED ID NOs: 9,10
 6. Claims: 1-3,7-16 partially idem for SED ID NOs: 11,12
 7. Claims: 1-3,7-16 partially idem for SED ID NOs: 13,14
 8. Claims: 1-3,7-16 partially idem for SED ID NOs: 15,16
 9. Claims: 1-3,7-16 partially idem for SED ID NOs: 17,18

BNSDOCID: <WO___9919493A3_IA>

INTERN ONAL SEARCH REPORT

mation on patent family members

ational Application No PCT/US 98/20807

Patent document cited in search report	nt į	Publication date	1	Patent family member(s)	Publication date		
WO 9103561	Α	21-03-1991	US	5212296 A	18-05-1993		
			ΑT	133201 T	15-02-1996		
•			AU	648036 B	14-04-1994		
			ΑU	6272990 A	08-04-1991		
			CA	2065439 A	12-03-1991		
	•		DE	69024979 D	29-02-1996		
			DE	69024979 T	17-10-1996		
			DK	554240 T	03-06-1996		
			EP	0554240 A	11-08-1993		
			ES	2082862 T	01-04-1996		
		•	JP	5500002 T	14-01-1993		
			US	5349127 A	20-09-1994		

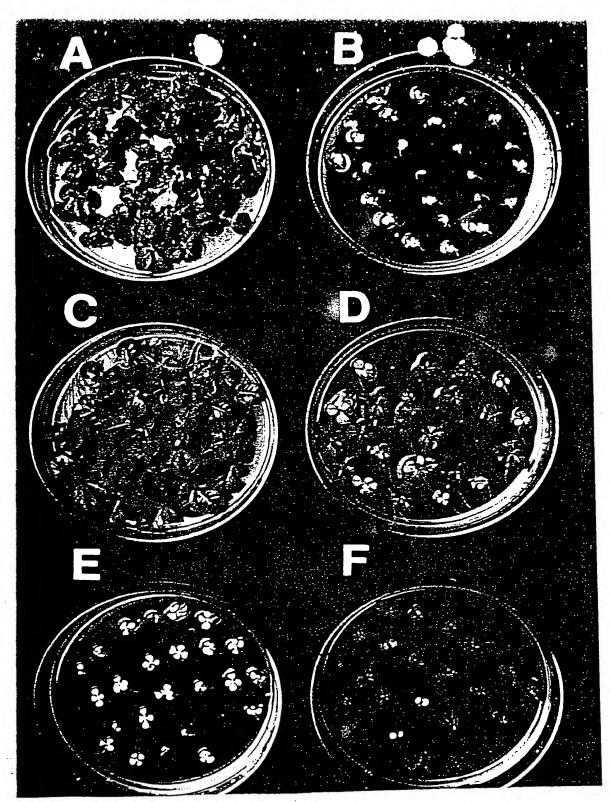


Figure 5

